

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
Белорусский государственный университет иностранных языков

ЯЗЫК В ЭПОХУ ЦИФРОВЫХ ТРАНСФОРМАЦИЙ И РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Сборник научных статей
по итогам II Международной научной конференции
Минск, 23–24 октября 2025 г.

Минск
БГУИЯ
2025

УДК 81'32/'33:004.8/.9(063)

ББК 81я4+95.4

Я41

Рекомендован Редакционным советом Белорусского государственного университета иностранных языков. Протокол 4/78 от 17.10.2025

Редакционная коллегия: Н. Е. Лаптева (*ответственный редактор*), О. В. Лушинская (*зам. ответственного редактора*), Т. М. Насонова, Е. Д. Долматова, О. М. Мижевич, Ю. В. Овсейчик, А. А. Соловей, Т. В. Аникеева, Ю. В. Дашкевич, А. С. Кочукова

Рецензенты: доктор филологических наук, профессор *Н. В. Чайка* (БГПУ им. М. Танка); кандидат филологических наук, доцент *Д. В. Степанова* (БГУИЯ)

Язык в эпоху цифровых трансформаций и развития искус-
Я41 ственного интеллекта : сб. науч. ст. по итогам II междунар. науч. конф., Минск, 23–24 окт. 2025 г. / редкол. : Н. Е. Лаптева (отв. ред.) [и др.]. – Минск : БГУИЯ, 2025. – 11,6 Мб.

ISBN 978-985-28-0324-3

Сборник научных статей подготовлен по итогам II Международной научной конференции «Язык в эпоху цифровых трансформаций и развития искусственного интеллекта», состоявшейся в БГУИЯ 23–24 октября 2025 г., по таким направлениям, как корпусные исследования языка и речи; технологии искусственного интеллекта для решения лингвистических задач; цифровая трансформация языкового образования; мультимодальная коммуникация в виртуальном пространстве; цифровые технологии в переводе и обучении переводу; междисциплинарные исследования на основе анализа текстовых данных и др.

Предназначен для исследователей в области экспериментальной и генеративной лингвистики, прикладной и компьютерной лингвистики, цифровой лингводидактики.

УДК 81'32/'33:004.8/.9(063)

ББК 81я4+95.4



Электронная версия издания
доступна в электронной библиотеке БГУИЯ
по ссылке e-lib.bsuf.by или по QR-коду

ISBN 978-985-28-0324-3

© УО «Белорусский государственный
университет иностранных языков», 2025

ПЛЕНАРНОЕ ЗАСЕДАНИЕ

УДК 81'42 + 811.929.1

Бредихин Сергей Николаевич
доктор филологических наук,
профессор
Северо-Кавказский федеральный
университет
г. Ставрополь, Россия

Sergey Bredikhin
Dr. habil. in Philology, Professor
North-Caucasus Federal University
Stavropol, Russia
bredichinsergey@yandex.ru

КОНВЕРГЕНЦИЯ КОДОВ В ВИРТУАЛЬНЫХ ДИСКУРСИВНЫХ ФОРМАЦИЯХ

В настоящем исследовании осуществлена попытка первичного описания принципов анализа контаминированного поликодового мультифокального стимул-реактивного пространства виртуального дискурса. Делимитированы ключевые характеристики иммерсивности и экстрагируемости генерализованного содержания на основе синергетического взаимодействия дифференциальных кодов. Определены возможности субординативного и координативного взаимодействия визуального, вербального и аудиального кодов. Постулируется преимущественно синестетическая представленность аудиального кода, формирующего общую тональность реминисценции. Базовый алгоритм анализа развернутого некооперативного взаимодействия адаптирован для исчисления суммарного иллюкутивно-перлокутивного соответствия конвергирующих кодов в сообщениях виртуального дискурса. Комплексный подход базируется на первичном автоматизированном анализе, включающем веб-скрейпинг, API, токенизацию и парсинг, дополняемый на финальной стадии семанто-синтаксическим праймированием и компаративным анализом лексико-семантических и структурно-топикальных соответствий исходного стимульного и реактивного комментирующего сообщений.

Ключевые слова: кодовая конвергенция; виртуальный дискурс; мультифокальность; иммерсивность; экстрагируемость; иконико-вербальная меметика.

CODE CONVERGATION WITHIN VIRTUAL DISCOURSE FORMATION

The study attempts to provide a primary description of the principles of analysis of contaminated polycode multifocal stimulus-reactive space of virtual discourse. The key characteristics of immersion and extractability of generalized content are delimited based on the synergetic interaction of differential codes. The possibilities of subordinate and coordinate interaction of visual, verbal and auditory codes are specified. The predominantly synesthetic representation of the auditory code, which forms the general tonality of reminiscence, is postulated. The basic algorithm for analyzing expanded non-cooperative interaction is adapted to calculate the total illocutionary-perlocutionary correspondence of converging codes in virtual discourse messages. The integrated approach is based on primary automated analysis, including web scraping, API, tokenization and parsing, supplemented at the final stage by semantic-syntactic priming and comparative analysis of lexical-semantic and structural-topical correspondences of the original stimulus and reactive commentary messages.

Key words: code convergence; virtual discourse; multifocality; immersion; extractability; iconic-verbal memetics.

Современное пространство интеракции характеризуется ростом роли различных медиаторов, в том числе технических, что переводит привычные дискурсотипы в сферу виртуального общения. Дополнительные функциональные ограничения, появляющиеся в условиях автоматизированной трансляции, а также снятие существующих в непосредственном взаимодействии барьеров формирует новые дискурсивные формации. При этом различные характеристики виртуальных медиа и формируемого на их основе виртуального медиадискурса приносят новые формы и возможности взаимодействия дифференциальных, часто не коррелирующих семиотических систем, способствующих реализации различных коммуникативных принципов, и не просто создающих параллельные экспликаторные ряды, но повышающих степень иллюкутивно-перлокутивного соответствия поликодового высказывания посредством синергетического эффекта.

Ключевыми характеристиками новых дискурсивных формаций виртуального пространства становятся иммерсивность и экстрагируемость, реализуемые в рамках виртуального пространства в виде различных интерпретативно-генеративных модусов, что закономерно ведет к мультифокальности деривации, рецепции, интерпретации генерализованного содержания.

Иммерсивность в новых виртуальных пространствах расширенного некооперативного взаимодействия достигается посредством конвергенции факторов «глубокой медиатизации» [8] как ухода инфоповода от необходимости эмпирической верификации для реализации максимального манипулятивного эффекта и возможности влиять на оценку описываемого в инициальных сообщениях события, то есть перманентной сменой статуса реципиента на продуцента, сокращением ситуации «даунтайма» в условиях необлигаторного, но открытого реагирования на инфоповод [3]. Следует также указать не только на качественные характеристики поликодовой репрезентации социально значимых событий, которые основаны на повышении уровня конвергенции бэкграунда как когнитивного, так и эмотивного активов активных агентов виртуального дискурса и пула реципиентов, потенциально заинтересованных в дальнейшей трансляции новых семантем и, таким образом, переходящих в статус агентов [4, с. 118], но и на количественную специфику частотной представленности тех или иных компонентов и их подкрепления в различных кодах. Иммерсивные характеристики в данном случае обеспечиваются квазиверификацией «истинности» инфоповода, определяемой и фиксируемой в поликодовом гипертекстуальном пространстве на основе частотности его представленности [9].

Характеристика экстрагируемости контента вне зависимости от степени его квазиистинностных компонентов обеспечивается в поликодовой виртуальной дискурсивной формации трилатеральным (мотивационным – инструментальным – результативным) полем формирования векторов создания [2, с. 108]. Вычленение значимостных компонентов детерминировано конститутивно, но объединение пертинентностных (мотивационных, актуальных, интерпретационных и социальных) областей, формирующих общую для

всего сетевого сообщества, ориентируемого на топикальную привязку инфоповода, релевантность [5], подкрепляется различными моделями некооперативного взаимодействия элементов различных кодов. Уровень экстрагируемости адекватных семантических и семиотических коррелятов ментальных областей (даже в рамках автоматизированного анализа креолизованных гипертекстовых пространств) существенно повышается при отходе от простого анализа однопорядковых элементов семиозиса, осуществляемого по формуле Бэкуса-Наура [1, с. 51], и привлечении к рассмотрению способов координативного или субординативного подкрепления по моделям $V_b + V_s + A = GC$ (V_b – вербальное инкодирование с вычленением топикальных токенов; V_s – визуальное инкодирование с семиотической иконикой; A – аудиальные элементы (чаще в синестетической аллюзивной репрезентации); GC – генерализованное содержание на основе поликодовой синергии).

Ситуативная и интерпретативная область детерминации контента сообщений виртуального медиадискурса базируется на доминировании вербальной текстовой составляющей, включающей идентификацию лексических (слова-маркеры), символических (математические, буквоподобные символы, технические знаки и т.п.), синтаксических (модели и синтаксические концепты)). Однако, если «ручная» дескрипция и определение семиотической топикальной нагрузки в данном случае не вызывает особых трудностей, то автоматизированная демаркация и типизация вызывает серьезные затруднения, поскольку требует использования нескольких программных продуктов (JAPE Transducer, ANNIE Gazetteer) [6], или разработки дополнительного обеспечения с возможностью интегрирования в корпусные менеджеры.

В случае распредмечивания компонентов визуального маркирования эмотивно-ассоциативных коннотаций в генерализованном смысловом содержании высказываний, которое также не является особой проблемой для семантизирующего сознания ввиду имманентной приданности данной семиотической системы и ее иконически-миметического характера (способности преломлять действительность в индивидуально-коллективном психоэмотивном опыте), осложнено для автоматизированной обработки по причине отсутствия возможности конвергентного анализа конситуативной семантики знаков в поликодовом единстве. Так, в настоящее время не существует автоматизированных систем вычленения саркастических, интенсифицирующих и иных контекстуальных иллокуций использования знаков эмотикона, разметки корпусов, содержащих данные символы, и уж тем более способов распредмечивания сложных изображений в виртуальном меметическом нарративе. Реминисценции и ассоциаты нового модифицированного прецедента базируются на ситуативных модальных и интерпретационных значимостях, получающих комплексное иконоково-вербальное меметическое подкрепление [7, с. 114].

Степень экстрагируемости прямо пропорциональна степени меметичности не только самого сообщения, но и ситуации мемезиса [10], как энклитичного и эклектичного механизма кодовой конвергенции.

Аудиальная импринтизация в подкреплении новых транслируемых в поликодовом виртуальном пространстве коннотатов задействует ввиду новых технических возможностей медиального персонально- и технико-опосредованного взаимодействия две формы – собственно фоновый аудиальный код и синестетический (вербальную аллюзию к аудиальному прецеденту). Вычленение собственно аудиальных семантем на настоящий момент возможно в топикальной детерминации и ситуативном ассоциировании только в «ручном» режиме, в то время как синестетика достаточно просто делимитируется на основании токенизации при интенциональном вводе ключевых лексем-вербализаторов.

Отметим, что последовательный учет и анализ разнокодовых импликаций не дает возможности делимитировать и адекватно интерпретировать все компоненты генерализованного содержания поликодового мультифокального высказывания виртуального дискурса, поскольку каждый из компонентов (от мемологемы, до вербального прецедента и ситуативной детерминанты) сам по себе обладает нарратогенным потенциалом.

$$\Gamma = \langle I, \{Si\} \ i \in I, \{Hi\} \ i \in I \rangle$$

I в данном случае представляет переменную, детерминируемую количеством семиотических систем прямой экспликации и маскировочной импринтизации, т.е. в ней учитываются как вербальные и визуальные компоненты, так и аудиальные и синестетические форманты, усиливающие генерализованную семантику или модифицирующие ее ситуативно; Si – является формой приматизации того или иного кода в определенной фокализации (агентной или реципиентной), при этом реализация модели с субординативной или координативной экспликацией в различных системах определяется в следующем подмножестве – $i \in I$, совокупность Hi позволяет в наиболее общем виде представить признаковый ряд компонентов импринтизации различных кодов, создающимся как результат деривативной поликодовой контаминации перманентно сменяемых иллокуций мультифокального виртуального пространства i , учитывающих смыслодеривационный и смысломодификационный потенциал ситуативно и контекстуально распрямечиваемых кодов в процессе мемезиса $S = \prod Si \ i \in I$ [1, с. 53].

Таким образом, основой распрямечивания генерализованного содержания комплексного поликодового мультифокального сообщения в виртуальных дискурсивных формациях предполагается доминирование эффекта семанто-семиотического прайминга [11]. Семиотический прайминг в данном случае включает в себя прогностику как иллокутивных интенционально формируемых областей эмоцио-, операцио-, когниогенеза текстового поликодового пространства, причем расчет должен производиться для каждого конкретного кода сначала сепаратно, а затем включаться в общее пространство конвергентного кодового взаимодействия как конситуативные и контекстуальные детерминанты.

Полностью автоматизированный анализ контаминированного поликодового мультифокального гипертекстового пространства виртуального дискурса с учетом кодовой конвергенции и различной степени иллюкутивно-перлокутивного соответствия каждого из них в настоящее время не представляется возможным. Однако применение имеющихся автоматизированных методик может быть модифицировано за счет комбинирования компьютерно-лингвистических методов автоматического отбора материалов исследования, а также методов, предоставляющих возможность дальнейшего углубленного анализа текстов. К таким методам можно отнести веб-скрейпинг (англ. *web scraping*) и последующий компьютерный синтаксический анализ (или парсинг, от англ. *parsing*). Веб-скрейпинг подразумевает извлечение информации с сайтов, включая текст, изображения и другие элементы, что особенно важно в полимодальном виртуальном медиапространстве. Оптимизировать работу веб-скрейпинга возможно за счет применения различных API (*application programming interface*), реализующих прямой доступ к веб-информации, распространяемой на интернетресурсах. Использование API позволяет избежать многих проблем, связанных с изменением структуры веб-страниц, поскольку материал предоставляется в заранее определенном формате, что обеспечивает более стабильный и структурированный доступ к данным и значительно ускоряет процесс извлечения информации. Последующий синтаксический анализ (парсинг) полученных веб-данных позволяет структурировать и отобрать необходимую для анализа информацию, а также задать первичный вид будущей базы данных исследования. Выделенные на основе автоматизированного анализа элементы различных кодов закономерно должны быть подвергнуты семанто-семиотическому праймированию для определения соответствия прогностики реальному перлокутивному эффекту, транслируемому в реактивном комментативном пространстве.

ЛИТЕРАТУРА

1. Бредихин С. Н. Теория некооперативных игр в порождении и интерпретации высказываний институционального дискурса // Профессиональная коммуникация: актуальные вопросы лингвистики и методики. 2023. № 16. С. 50–58.
2. Бредихин С. Н., Каменский М. В., Шибкова О. С. Новые медиа в процессе модификации концептуально-валерного пространства // Вестник Воронежского государственного университета. Серия: Филология. Журналистика. 2024. № 2. С. 107–111.
3. Бредихин С. Н., Пелевина Н. А. Сентимент-анализ интенсификации интимизиционных компонентов комментативного текста в фейковом дискурсе // Изоморфные и алломорфные признаки языковых систем. 2025. № 7. С. 55–63.
4. Бредихин С. Н., Шибкова О. С., Гусаренко С. В. Когнитивный блендинг как основа создания релевантности высказывания // Актуальные проблемы филологии и педагогической лингвистики. 2023. № 4. С. 118–128.
5. Бредихина Ю. И. Пространство целеполагания как форма учета ценностных ориентаций в дискурсе социальной работы (пертинентность vs релевантность) // Гуманитарные и юридические исследования. 2024. Т. 11, № 1. С. 168–174.

6. Каменский М. В., Бредихин С. Н. Алгоритмические процедуры идентификации рекламных текстов в дискурсивном пространстве средств массовой информации // Вестник Волгоградского государственного университета. Серия 2: Языкознание. 2025. Т. 24, № 1. С. 64–78.
7. Марченко Т. В., Бредихин С. Н., Лепилкина О. И. Специфика формирования ассоциативного потенциала нового прецедентного феномена и поликодовой объективации в современном британском медиадискурсе // Вестник Калмыцкого университета. 2023. № 3(59). С. 109–115.
8. Couldry N., Hepp A. The continuing lure of the mediated centre in times of deep mediatization: Media Events and its enduring legacy // Media, Culture & Society. 2018. № 40 (1). P. 114–117.
9. Hepp A., Hjarvard S., Lundby K. Mediatization: theorizing the interplay between media, culture and society // Media, Culture & Society. 2015. № 37 (2). P. 314–324.
10. Tyler T. Memetics: Memes and the Science of Cultural Evolution. New York : Mersenne Publishing, 2011. 326 p.
11. Weingarten E., Chen Q., McAdams M. Yi J., Hepler J., Albarracín D. From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words // Psychological Bulletin. 2016. № 142 (5). P. 472–497.

УДК 004.91+372.881.111.1

Горожанов Алексей Иванович

доктор филологических наук,
профессор кафедры грамматики
и истории немецкого языка
факультета немецкого языка
Московский государственный
лингвистический университет
г. Москва, Россия

Alexey Gorozhanov

Doctor of Science (Philology),
Professor in the Department of German
Language Grammar and History
Moscow State Linguistic University
Russian Federation
a_gorozhanov@mail.ru

Гусейнова Иннара Алиевна

доктор филологических наук,
профессор кафедры лексикологии
и стилистики немецкого языка
факультета немецкого языка
Московский государственный
лингвистический университет
г. Москва, Россия

Innara Guseynova

Doctor of Science (Philology),
Professor in the Department of German
Language Lexicology and Stylistics
Moscow State Linguistic University
Russian Federation
ginnap@mail.ru

**СОСТАВЛЕНИЕ ДИДАКТИЧЕСКИХ МАТЕРИАЛОВ ДЛЯ ОБУЧЕНИЯ
АКАДЕМИЧЕСКОМУ ПИСЬМУ НА ОСНОВЕ КОРПУСНЫХ ДАННЫХ**

Исследование направлено на то, чтобы показать возможности автоматической и автоматизированной обработки аутентичных текстов (176 научных статей международного журнала “Frontiers in Education”) с целью составления дидактических материалов для обучения академическому письму на английском языке. В качестве технического инструмента используется авторский программный комплекс «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» (свидетельство Роспатент № 2023683209). Путем различных запросов в корпусном менеджере авторы демонстрируют процедуру извлечения аутентичных контекстов, которые содержат частотные сочетания токенов длиной 3–6 единиц, две единицы, а также типичные для научной статьи отдельные лексемы. В результате делается вывод об эффективности корпусного подхода и используемого программного комплекса для составления дидактических материалов по академическому письму. Подчеркивается важность высокого качества исходного корпуса текстов, от которого зависит достоверность получаемых материалов в плане их реальной ценности для образовательного процесса.

Ключевые слова: академическое письмо; дидактические материалы; научный дискурс; исследовательская статья; корпусная лингвистика; корпусный менеджер.

**COMPILATION OF DIDACTIC MATERIALS
FOR TEACHING ACADEMIC WRITING BASED ON CORPUS DATA**

The study aims to demonstrate the capabilities of automatic and automated processing of authentic texts (176 research articles of the international journal “Frontiers in Education”) in order to compile didactic materials for teaching academic writing in English. The authors’ software package “Balanced Linguistic Corpus Generator and Corpus Manager” (Rospatent

Certificate No. 2023683209) is applied as a technical tool. By means of various queries through the corpus manager, the authors demonstrate the procedure for extracting authentic contexts that contain frequent combinations of tokens 3-6 units long, two units long, as well as single lexemes typical for a research article. As a result, a conclusion is made about the effectiveness of the corpus approach and the software package used for compiling didactic materials on academic writing. The importance of the high quality of the text corpus, on which the reliability of the obtained materials depends in terms of their real value for the educational process, is emphasized.

Key words: academic writing; didactic materials; scientific discourse; research article; corpus linguistics; corpus manager.

Предлагаемое исследование направлено на то, чтобы показать возможности автоматической и автоматизированной обработки аутентичных текстов (научных статей) с целью составления дидактических материалов для обучения академическому письму на английском языке. При этом фокус делается на получении частотных слов и словосочетаний, типичных для жанра исследовательской статьи научного дискурса.

Языковым материалом работы является сбалансированный лингвистический корпус текстов 176 научных статей из журнала “Frontiers in Education” объемом 16 233 предложения или 483 117 токенов, который был собран в 2024–2025 гг. Этот журнал избран в качестве образца как международное рецензируемое издание, индексируемое в базе данных “Scopus” (Q2).

В качестве технического инструмента используется авторский программный комплекс «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» (свидетельство Роспатент № 2023683209), включая дополнения: «Модуль расширения базы данных сбалансированного лингвистического корпуса коннотационными значениями» (свидетельство Роспатент № 2024686126), «Модуль поиска фразеологизмов в сбалансированном лингвистическом корпусе» (свидетельство Роспатент № 2024686050), «Модуль оценки контекста лексической единицы» (свидетельство Роспатент № 2024685229) [2]. Привлечение современного корпусного инструментария формирует новизну нашей работы.

Актуальность исследования обусловлена высоким интересом со стороны научно-педагогического сообщества как к проблеме обучения академическому письму [1; 6], в том числе, преподавателей вузов [4], так и к лингвистическим аспектам составления научного текста [3; 7].

Здесь мы затрагиваем процесс подготовки дидактических материалов только в части получения образцов-предложений, демонстрирующих употребление того или иного языкового явления в письменном научном тексте.

Прежде всего, мы произвели запрос к корпусу, результатом которого явились списки частотных словосочетаний длиной от трех до шести токенов, которые встречаются в статьях 20 и более раз. При этом под токенами понимались не только слова, но и знаки пунктуации, числа и символы (% , # , @ и пр.). Для последовательности из трех токенов список составил 694 пунк-

та. С увеличением длины последовательности он закономерно уменьшался и для шести токенов сократился до четырех пунктов, касающихся, по сути, одного словосочетания:

- 1) the perceived utility of AI in;
- 2) the perceived utility of AI integration;
- 3) perceived utility of AI integration in;
- 4) utility of AI integration in education.

Этот результат позволил далее произвести запрос к текстам предложений и выявить полные контексты, содержащие фрагмент “the perceived utility of AI”, например:

[8147]¹: **The perceived utility of AI** integration within education field is positively sustained by the ongoing support and available resources that can be used in order to ensure a sustaining teachers’ use of AI tools. – [8147]: **Ощущаемая полезность** интеграции **ИИ** в сферу образования положительно подкрепляется постоянной поддержкой и доступными ресурсами, которые можно использовать для обеспечения постоянного использования преподавателями инструментов **ИИ**².

[8246]: Results and ideas presented so far make clear the importance that different practical modalities to use AI within the teaching process have an important role in augmenting **the perceived utility of AI** integration in education. – Представленные результаты и идеи наглядно демонстрируют важность того, что различные практические способы использования **ИИ** в образовательном процессе играют значительную роль в повышении **ощущаемой полезности** интеграции **ИИ** в сферу образования.

[8447]: Moreover, the study suggests that addressing these challenges directly – such as by simplifying AI tools, providing more intuitive interfaces, and ensuring that AI enhances rather than detracts from teacher-student interactions – can significantly improve **the perceived utility of AI** in education. – Более того, наше исследование показывает, что решение этих проблем напрямую – например, путем упрощения инструментов **ИИ**, предоставления более интуитивно понятных интерфейсов и обеспечения того, чтобы **ИИ** улучшал взаимодействие обучающего и обучающегося, а не мешал ему – может значительно повысить **ощущаемую полезность ИИ** в сфере образования.

Приведенные выше образцы возможно использовать как для демонстрации контекстного окружения словосочетания, так и для формирования переводческих умений.

Анализ списка сочетаний из пяти токенов позволил выявить такие конструкции, как “on the one / other hand”, “as well as the”, “it is important to”, “attitudes toward parental involvement”, “in the private / public sector”, “to adopt and use AI”, “it is necessary to”, “the fourth industrial revolution”.

¹ Дается порядковый номер предложения в корпусе.

² Здесь и далее перевод наш – А. Г., И. Г.

Применяя таким же образом запрос к текстам предложений, оператор корпусного менеджера может получить полные контексты, которые, при необходимости, могут выходить за рамки одного предложения, например:

[6435]: Table 2 approves hypotheses 1 to 3, showing a positive correlation between teachers' attitudes toward parental involvement and transformational, authentic, and participative leadership styles. – [6435]: Таблица 2 подтверждает гипотезы 1–3, показывая положительную корреляцию между отношением преподавателей к участию родителей в образовательном процессе и трансформационным, аутентичным и партисипативным стилями лидерства.

[3306]: Furthermore, the challenges of the fourth industrial revolution can be overcome with strategic, systemic plans and collaboration with industries. – [3306]: Более того, вызовы четвертой промышленной революции можно преодолеть с помощью стратегических системных планов и сотрудничества с представителями промышленности.

Еще одной, автоматизированной, процедурой является анализ сформированного автоматически алфавитного списка предложений корпуса, который позволяет достаточно быстро выявить типичные для научного текста сочетания из двух токенов или даже единичные лексемы, например:

[5794]: **According to the literature**, this dimension “enhances students' well-being and learning motivation” (Burić and Kim, 2020, p. 5; Praetorius et al., 2018). – [5794]: **Согласно научным источникам**, это измерение «улучшает благополучие студентов и мотивацию к обучению» (Burić and Kim, 2020, p. 5; Praetorius et al., 2018).

[6943]: **Additionally**, 16 % of students anticipated the development of more realistic and effective simulation software to further enhance the course's engagement. – [6943]: **Кроме того**, 16 % студентов ожидали разработки более реалистичного и эффективного программного симулятора, что еще больше повысит мотивацию к изучению дисциплины.

[8945]: **Another notable concern is** the increased sense of isolation or disengagement in online learning environments (Kaufmann and Vallade, 2020). – [8945]: **Еще одной заметной проблемой** является возросшее чувство изоляции или отчужденности в среде онлайн-обучения (Kaufmann and Vallade, 2020).

Заметим, что мы ограничиваемся только лишь частотными словосочетаниями или отдельными словами, не затрагивая такие важные аспекты подъязыка (см. подробнее [5]), как терминологию, особенности употребления грамматических конструкций и пр. Однако продемонстрированных возможностей достаточно, чтобы сделать вывод об эффективности корпусного подхода и используемого программного комплекса для составления дидактических материалов по академическому письму. При этом ключевую роль играет процедура сборки лингвистического корпуса, от качества которого будет зависеть достоверность получаемых материалов в плане их реальной ценности для образовательного процесса.

Перспективой работы может быть усовершенствование программного обеспечения с целью перевода автоматизированных процедур в разряд автоматических, например, разработка функции получения частотного списка первых (N-ных) токенов предложений корпуса.

ЛИТЕРАТУРА

1. Васильева Н. Л. Жанровая педагогика в обучении академическому письму / Н. Л. Васильева, Е. В. Алешинская // Филологический класс. – 2024. – Т. 29, № 1. – С. 143–151. – DOI 10.26170/2071-2405-2024-29-1-143-151. – EDN UBSZZO.
2. Горожанов А. И. Архитектура сбалансированного лингвистического корпуса, полученного автоматическим путем (опыт Московского государственного лингвистического университета) / А. И. Горожанов // Вестник Московского государственного лингвистического университета. Гуманитарные науки. – 2024. – № 11(892). – С. 24–30. – EDN BCSCXQ.
3. Зашихина И. М. Формат научных публикаций IMRaD для социально-гуманитарных исследований: шанс быть услышанным / И. М. Зашихина, О. В. Печинкина // Высшее образование в России. – 2022. – Т. 31, № 10. – С. 150–168. – DOI 10.31992/0869-3617-2022-31-10-150-168. – EDN JZRZLD.
4. Научный потенциал личности: развитие профессионально значимых компетенций в вузе / А. К. Арынгазин, Г. Ж. Фахрутдинова, А. А. Арынгазин, А. А. Арынгазин // Вестник Томского государственного университета. – 2023. – № 496. – С. 5–12. – DOI 10.17223/15617793/496/1. – EDN NMSIJE.
5. Писарик О. И. Доля английского профессионального подязыка в текстах учебных материалов для студентов строительных специальностей / О. И. Писарик, А. И. Горожанов // Филологические науки. Вопросы теории и практики. – 2021. – Т. 14, № 3. – С. 928–932. – DOI 10.30853/phil210082. – EDN CIRARX.
6. Boginskaya O. A. A Comparison of Explicit and Implicit Approaches to EAP Teaching to Postgraduate Students / O. A. Boginskaya // Higher Education in Russia. – 2024. – Vol. 33, No. 2. – P. 148-161. – DOI 10.31992/0869-3617-2024-33-2-148-161. – EDN SJURSZ.
7. Tikhonova E. V. Wordiness in academic writing: a systematic scoping review / E. V. Tikhonova, D. A. Mezentseva // Research Result. Theoretical and Applied Linguistics. – 2024. – Vol. 10, No. 1. – P. 133-157. – DOI 10.18413/2313-8912-2024-10-1-0-8. – EDN WERICN.

УДК 81-11

Евдокимова Вера Вячеславовна

кандидат филологических наук,
доцент
СПбГУ
г. Санкт-Петербург, Россия

Vera Evdokimova

PhD in Philology,
Associate Professor
Saint Petersburg State University
Saint-Petersburg, Russia
v.evdokimova@spbu.ru

Скрелин Павел Анатольевич

доктор филологических наук,
профессор
СПбГУ
г. Санкт-Петербург, Россия

Pavel Skrelin

Doctor of Science in Philology,
Professor
Saint Petersburg State University
Saint-Petersburg, Russia
p.skrelin@spbu.ru

Евграфова Карина Владимировна

кандидат филологических наук,
доцент
СПбГУ
г. Санкт-Петербург, Россия

Karina Evgrafova

PhD in Philology,
Associate Professor
Saint Petersburg State University
Saint-Petersburg, Russia
k.evgrafova@spbu.ru

ФОНЕТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ПСИХОФИЗИОЛОГИЧЕСКОГО СОСТОЯНИЯ ЧЕЛОВЕКА

В статье приводится краткий обзор работ, которые проводятся на кафедре фонетики и методики преподавания иностранных языков СПбГУ в рамках исследований физического, психофизиологического и эмоционального состояния человека по речевому сигналу. Описаны основные направления работ и продолжающиеся исследования. Дистанционная оценка психофизиологического и эмоционального состояния оператора, и верификация принимаемого речевого сообщения, и формализация тонких фонетических признаков диктора для качественного синтеза, распознавания его речи стала возможной в связи с развитием вычислительных возможностей и систем искусственного интеллекта. Среди направлений текущих исследований кафедры описаны работы по голосовой усталости и патологиях, когнитивной нагрузке, когнитивному, физическому и эмоциональному состояниям. В целом, перспективными являются и работы с применением систем искусственного интеллекта, и использование непосредственно фонетических признаков для получения результатов высокого уровня эффективности.

Ключевые слова: фонетика; акустический анализ речи; психофизиологическое состояние; когнитивные нарушения; речевые расстройства.

PHONETIC CHARACTERISTICS OF THE PSYCHOPHYSIOLOGICAL STATE OF A PERSON

The article provides a brief overview of the work that is being carried out at the Department of Phonetics and Methods of Teaching Foreign Languages of St. Petersburg State

University in the framework of work on determining the physical, psychophysiological and emotional state of a person by speech and voice. The main areas of work and ongoing research are described. Remote assessment of the psychophysiological and emotional state of the operator, and verification of the received speech message, and formalization of subtle phonetic features of the speaker for high-quality synthesis, recognition of his speech became possible due to the development of computing capabilities and artificial intelligence systems. Among the areas of current work at the department, studies of voice fatigue and pathologies, cognitive load, cognitive states, physical and emotional state are described. In general, both work using artificial intelligence systems and the use of direct phonetic features to obtain high-level results are promising.

Key words: phonetics; speech acoustics; psychophysiological state; cognitive impairment; speech disorders.

Человеческая речь – это сложнейшее явление. Она отражает мышление человека и прежде всего способствует информационному обмену между людьми. Каждый элемент речевого потока несет слушателю разнообразную информацию. Однако речь – не единственный информационный речевой канал. Важна не только информационная составляющая сообщения, но и как человек говорит. Это отражает его психофизиологическое состояние, его вольное или невольное стремление передать свое отношение к содержанию высказывания, к собеседнику к окружающей обстановке, к ситуации общения и т. д. Если языковые единицы передают содержание сообщения, то его фонетическая картина предельно сложна и неоднозначна [1].

Текущий уровень технического прогресса в области обработки речи делает возможным не только распознавать смысл речевого сообщения, но и анализировать фонетические признаки. Появление новых вычислительных возможностей позволило ставить новые прикладные задачи. Это и дистанционная оценка психофизиологического и эмоционального состояния оператора, и верификация принимаемого речевого сообщения, и формализация тонких фонетических признаков диктора для качественного синтеза, распознавания его речи [2].

Зафиксированная методами современной звукозаписи речь представляет собой исходный информационный массив, в котором заложена информация обо всех ее особенностях. Их выявление зависит от выбранного алгоритма обработки собранного материала. Адекватность и ценность результатов обработки определяется всем процессом работы с речевым материалом. На этапе сбора материала нужно обеспечить его достаточность и соответствие поставленной задаче, на этапе подготовки выделить и сформировать аудиофайлы, пригодные для дальнейшей работы, на этапе обработки правильно выбрать метод исследования.

Устойчиво выделить характерные просодические и экстралингвистические особенности речи диктора можно только на основе обработки обширного речевого материала. Современные звукозаписывающие системы дают возможность сбора и стабильного хранения огромных речевых массивов. Однако в общем случае такие массивы непригодны для фонетического

анализа. Они могут представлять собой многосторонние диалоги, сопровождающиеся к тому же музыкой, производственными, транспортными и прочими шумами. Необходима процедура выделения речевых отрезков для решения поставленных фонетических задач распознавания [2].

Внедрение речевых технологий в построение новых интерактивных систем потребовало сосредоточить внимание не только на лингвистическом содержании речевой посылки, но и на его эмоциональной окраске как важном дополнительном канале передачи информации [3].

Анализ особенностей диалога в системах «человек-человек» и «человек-машина» позволяет выделить задачу учета эмоциональной и психофизиологической стороны речи и, соответственно, основные направления исследований.

Если говорить о распознавании эмоционального состояния диктора, то современная вычислительная техника открывает возможность постепенного набора информации об эмоциях и их классификации. Математический аппарат определения места эмоции в n -мерном пространстве может быть различным, но имеет общее целевое назначение [2]. Распознавание эмоций становится важной составной частью программного обеспечения центров сервисного обслуживания, систем жизнеобеспечения, систем обучения, интерактивных игр и т. д. Регистрация эмоций пользователя, таких как гнев, замешательство, страх, смущение, дает важнейшую информацию для улучшения качества обработки речевого сообщения, понимания запросов пользователя и, в конечном счете, для принятия правильного решения [3]. Часто оценка эмоций позволяет определить проблемное место в беседе, изменить стратегию диалога, заметить негативную реакцию собеседника [3].

В приложении к компьютеру изучение проявления в речи человеческих эмоций открывает возможности их моделирования и внедрения, например, при разработке эмоционально окрашенной речи для дружественных интерфейсов. Существуют программы, уже дополненные технологиями такого рода. Они обеспечивают естественность коммуникации между человеком и машиной [4, 5, 6].

Вопрос оценки физического состояния оператора по его речи лежит в области, пограничной с оценкой эмоционального состояния, и неразделим с ним. Большинство исследователей отмечают взаимопроникновение пространств признаков этих направлений, но предпочитают проводить методическую проработку на базе хорошо экспериментально классифицируемых эмоций. В то же время отмечается необходимость решения задач объективной оценки физического состояния оператора (безопасность на транспорте, авиация, медицина). Разрабатываемый при оценке эмоционального состояния математический аппарат может быть, при необходимости, переориентирован без значительных изменений на оценку физического состояния. Можно предполагать, что наибольшую трудоемкость представит сбор и классификация речевого материала.

На кафедре фонетики и методики преподавания иностранных языков СПбГУ продолжаются научные работы по анализу речи, которые могут стать основой новых способов оценки психофизиологического состояния человека [7, 8].

Продолжаются исследования голосовой усталости и патологий голоса совместно с фониатром Мариинского театра Н. В. Швалевым [9–11]. Голосовая усталость является сложным физиологическим явлением, которое проявляется в том числе и в изменении акустических параметров голоса. Его исследование представляет интерес как для уточнения моделей порождения голоса, так и для разработки методов объективной оценки качества голоса. Наиболее сильное проявление состояния голосовой усталости можно наблюдать у профессионалов голоса: преподавателей, певцов, актеров, гидов, переводчиков и т. д.

Для представителей голосо-речевых профессий особую важность представляет точная и ранняя диагностика нарушений работы голосового аппарата. Наряду с существующими инвазивными методами обследования, акустический анализ находит широкое практическое применение для безопасной и удобной оценки голосовой функции, обнаружения голосовых расстройств, а также эстетической оценки голоса. В выпускной квалификационной работе Софьи Куряковой «Проявление в голосе признаков заболевания речевого тракта» (2024 г.) особое внимание уделяется обзору существующих акустических коррелятов патологий голосового аппарата, а также анализу методов для их детектирования. Особенность данного исследования обусловлена ее направленностью на анализ вокальной и невокальной речи путем сравнения параметров джиттера, шиммера, HNR, а также спектральных характеристик при наличии или отсутствии патологии. Результаты исследования продемонстрировали существенные отличия акустических характеристик вокальной и невокальной речи в случае здорового или патологического состояния голосового аппарата. Кроме того, были отмечены значимые тенденции в связи с выбором регистра в пении (средний, высокий и низкий). Оценка результатов показала, что при фонации в высоком регистре отмечается уменьшение параметров джиттера и шиммера, увеличение HNR. Для анализа вокальной речи при мониторинге голосового здоровья был обоснован выбор в пользу спектрального анализа, а также параметров, рассчитывающих долю шумового и гармонического компонента в сигнале. Для групп пациентов были выявлены наиболее значимые параметры в зависимости от клинической картины.

В работе К. В. Евграфовой, Н. В. Швалева и Н. С. Соколовой при изучении акустических характеристик голоса и речи преподавателей высшей школы в состоянии утомления до начала пандемии были получены следующие результаты: изменения значений частоты основного тона, диапазона частоты основного тона, средних длительностей гласных и согласных, изменение количества и длительности пауз, общее сокращение интенсивности артикуляции согласных и изменения в месте образования ряда согласных

[12]. В цели исследования, проведенного в течение учебных семестров 2020–2021 гг., входил сравнительный анализ голосовой нагрузки преподавателей нашего университета, вызванной необычными условиями работы в период пандемии COVID-19.

В 2023–2025 годах проходили работы по определению фонетических характеристик как маркеров когнитивной нагрузки под руководством В. В. Евдокимовой была защищена магистерская диссертация М. Р. Максимовой [13]. Цель работы заключалась в изучении фонетических особенности речи при когнитивной нагрузке. Для выполнения работы необходимо решить следующие задачи: сравнение характеристик речи в различных условиях: при когнитивной нагрузке и при ее отсутствии; при когнитивной нагрузке и при предпаузальном удлинении, изучение визуальных маркеров когнитивной нагрузки и их связи с проявлениями когнитивной нагрузки в голосе. Были сделаны выводы о том, что при когнитивной нагрузке скорость артикуляции, в целом, понижается. Человек меняет также и интонационные характеристики речи, более часто использует ровные интонационные контуры.

В большинстве случаев визуальные и фонетические признаки когнитивной нагрузки совпадали по времени начала. Сегменты с одновременным проявлением мимических и речевых показателей характеризовались большей длительностью по сравнению с участками, в которых наблюдались только визуальные признаки. Это может означать, что распознавание когнитивной нагрузки одновременно по характеристикам речи и мимики способно дать более точный результат по сравнению с распознаванием только по акустическим параметрам или только по мимическим признакам.

На кафедре фонетики и методики преподавания иностранных языков ведутся работы по работе с фонетическими признаками речевых расстройств вместе с логопедами из Ярославского государственного педагогического университета им. Ушинского – был составлен фонетически сбалансированный тестовый материал для автоматического анализа артикуляционных нарушений.

Также на кафедре ведутся работы по определению фонетических характеристик для распознавания когнитивных нарушений по типу Альцгеймера. Фонетические признаки при традиционном анализе дают более высокую эффективность, чем системы, основанные на использовании технологий искусственного интеллекта [14]. Анализ артикуляции вносит наибольший вклад и в определение характера и других нейро-дегенеративных заболеваний (Паркинсон и умственные расстройства) [15].

Таким образом, можно констатировать, что методы и инструменты анализа речевого сигнала, хорошо известные фонетистам по описаниям артикуляторных, акустических и перцептивных свойств произносительной нормы языка, его социальных и интерферируемых вариантов, имеют широкие перспективы применения в исследованиях и моделировании самых разных аспектов речевого поведения человека.

ЛИТЕРАТУРА

1. Евдокимова В. В. Акустический анализ речевого сигнала., Санкт-Петербург: Скифия-принт, 2023. 128 с.
2. Евдокимова, В. В. Фонетические методы распознавания психофизиологического состояния человека. Санкт-Петербург: Скифия-принт, 2023. 108 с.
3. А. А. Поволоцкая, В. В. Евдокимова, П. А. Скрелин. Запись и апробация набора речевых данных для распознавания негативных эмоций в речи // Terra Linguistica. – 2023. – Т. 14, № 2. – С. 59–76.
4. Emotional Speech Recognition of Holocaust Survivors with Deep Neural Network Models for Russian Language. Bukreeva, L., Guseva, D., Dolgushin, M., Evdokimova, V., Obotnina, V. (2023). In: Karpov, A., Samudravijaya, K., Deepak, K.T., Hegde, R.M., Agrawal, S.S., Prasanna, S.R.M. (eds) Speech and Computer. SPECOM 2023. Lecture Notes in Computer Science(), vol 14338. Springer, Cham.
5. Developing a Question Answering System on the Material of Holocaust Survivors' Testimonies in Russian. Bukreeva, L., Guseva, D., Dolgushin, M., Evdokimova, V., Obotnina, V. (2023). In: Karpov, A., Samudravijaya, K., Deepak, K. T., Hegde, R. M., Agrawal, S. S., Prasanna, S. R. M. (eds) Speech and Computer. SPECOM 2023. Lecture Notes in Computer Science(), vol 14339. Springer, Cham.
6. Автоматический анализ эмоциональной мимики и интонации в русской речи. Щербаков, П. П. Кочеткова, У. Е. Скрелин, П. А. Борисов, Н. В. Федькин, П. С. Евдокимова, В. В. & Долгушин, М. Д. 2024, Язык-Музыка-Жест: информационные перекрестки (LMGIC-2022): МАТЕРИАЛЫ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ Санкт-Петербург, 18–20 апреля 2024 г. Эйсмонт, П. М. & Алексеева-Нилова, Т. Е. (eds.). Издательство Санкт-Петербургского университета, с. 102–104.
7. THE EFFECT OF PHYSICAL ACTIVITY ON THE PHONETIC CHARACTERISTICS OF SPEECH. Evdokimova, V. V. & Zakharchenko, E. A., 2019, Models and Analysis of Vocal Emissions for Biomedical Applications: 11th International Workshop, December, 17–19, 2019. Manfredi, C. (ed.). Firenze: Firenze University Press, p. 237–240 4 p. (Proceedings e report; vol. 122).
8. Phonetic changes in speech in various physical states / A. D. Ananeva, D. S. Gineva, V. V. Evdokimova [et al.] // Proceedings of the 6th Saint Petersburg Winter Workshop on Experimental Studies of Speech and Language (Night Whites 2022) : Materials of the symposium, Saint Petersburg, 15–16 декабря 2022 года. – Saint Petersburg: Skifia-print, 2023. – P. 14–15.
9. Acoustic analysis of vocal fatigue in professional voice users. Evgrafova, K. V. & Evdokimova, V. V., 2011, Models and Analysis of Vocal Emissions for Biomedical Applications – 7th International Workshop, MAVEBA 2011. Manfredi, C. (ed.). Firenze University Press, p. 153–156.
10. Detection of the Frequency Characteristics of the Articulation System with the Use of Voice Source Signal Recording Method. Evdokimova, V., Evgrafova, K.,

Skrelin, P., Chukaeva, T., Shvaley, N. (2013). In: Železný, M., Habernal, I., Ronzhin, A. (eds) *Speech and Computer. SPECOM 2013. Lecture Notes in Computer Science()*, vol 8113. Springer, Cham.

11. Речевой корпус для исследования голосовой усталости, связанной с профессиональной нагрузкой. Евграфова, К. В., Евдокимова, В. В., Скрелин, П. А. & Чукаева, Т. В., 2017, Анализ разговорной русской речи (АРЗ – 2017): Труды седьмого междисциплинарного семинара. Политехника-принт, с. 26–31.

12. К. В. Евграфова, Н. С. Соколова, Н. В. Швалева. Особенности артикуляции и голоса преподавателей высшей школы в процессе онлайн-обучения // *Современные векторы развития специального и инклюзивного образования: Сборник научных статей II международной научно-практической конференции, Ярославль, 13–14 февраля 2024 года.* – Ярославль: Ярославский государственный педагогический университет им. К. Д. Ушинского, 2024. – С. 199–205.

13. Максимова, М. Р. Фонетические характеристики речи при когнитивной нагрузке / М. Р. Максимова // *Фонетический лицей: Сборник статей.* – Санкт-Петербург : ООО "Скифия-принт", 2024. – С. 44–48.

14. Costantini, G., Cesarini, V., Di Leo, P., Amato, F., Suppa, A., Asci, F., Pisani, A., Calculi, A., Saggio, G.: Artificial intelligence-based voice assessment of patients with parkinson's disease off and on treatment: machine vs. deep-learning comparison. *Sensors* 23(4), 2293 (2023).

15. Garcia, A.M., Arias-Vergara, T., C Vasquez-Correa, J., Noth, E., Schuster, M., Welch, A. E., Bocanegra, Y., Baena, A., Orozco-Arroyave, J. R.: Cognitive determinants of dysarthria in parkinson's disease: an automated machine learning approach. *Movement Disorders* 36(12), 2862-2873 (2021).

УДК 81-139

Сулейманова Ольга Аркадьевнадоктор филологических наук,
профессор
Московский городской
педагогический университет
г. Москва, Россия**Olga Suleimanova**PhD, Professor
Moscow City University
Moscow, Russia
souleimanovaoa@mgpu.ru

ПРАГМАТИКА И ДИСКУРСИВНАЯ ЛИНГВИСТИКА: ПРИНЦИП ДОПОЛНИТЕЛЬНОСТИ?

В статье рассматривается корреляция двух ключевых лингвистических парадигм: прагматики и дискурсивной лингвистики. Актуальность работы определяется необходимостью провести границу между ними с точки зрения различий в познавательных установках, материала исследования и методики его проведения. Методика исследования базируется на анализе высказываний, в рамках которых обосновывается и проводится разграничение прагматического и дискурсивного содержания. Материалом работы послужили высказывания преимущественно театрального медиадискурса, содержащие пословицы, аллюзии и различную энциклопедическую информацию. Исследование позволило провести границу между дискурсивным знанием, понимаемым как «невыраженное», не эксплицированное знание, выводимое из широкого контекста, тогда как прагматическое содержание интерпретируется как в принципе логично выводимое знание из семантического содержания высказывания.

Ключевые слова: дискурсивный; прагматика; парадигма в лингвистике; методика дискурсивного анализа.

The paper focuses on two domineering linguistic paradigms, i. e. pragmatics and discourse analysis, the borderline between them being a matter of hot disputing in modern linguistics as regards their cognitive potential, the empirical base and research techniques. The research procedure relies on analyzing the text samplings where the discourse and pragmatic contents are exemplified, on distinguishing the contextual pragmatic information from the discourse implicit content that must be elicited (deduced / inferred) from the shared knowledge of the communicants. The sampling contains the theatrical media discourse sentences featuring allusions and other inter discourse elements, shared culture-related knowledge which is not explicit. The pragmatic information refers to the content which can be logically inferred in the form of presupposition, or implications from the text and which does not necessarily call for any “cultural” (in the most broad sense of the word) decoding.

Key words: analysis of discourse; pragmatics; linguistic paradigm; distinguishing pragmatics from discourse analysis.

На определенных этапах развития лингвистической мысли происходит накопление методологических парадигмально обусловленных языковых построений, которые часто пересекаются и накладываются друг на друга [1; 2], что определяет необходимость установления корреляции между существующими разрабатываемыми парадигмами и определяет **актуальность** таких попыток, равно как и интерес лингвистов к выработке непротиворечивых методологических положений. В данной работе принимается попытка наметить демаркационную линию между двумя

господствующими парадигмами (см. о парадигмальной корреляции, например, в исследовании С. А. Потякайло, в котором на основе анализа всех российских диссертационных работ по пяти специальностям за 2000–2023 гг. показана их ведущая роль в лингвистических исследованиях) – прагматикой и дискурсивным анализом в той его разновидности, что соответствует конвексационному анализу в американской лингвистической традиции, или дискурсивному анализу, разрабатываемому в большинстве работ в отечественной лингвистической традиции (см, например [3]). Отметим вместе с тем, что предлагаемая демаркация в принципе также валидна для направления критического анализа дискурса, поскольку проводится на тех же основаниях, что и иные типы дискурсивных направлений.

Принципиальными для современной интерпретации дискурсивного анализа представляются следующие исходные положения, а именно: понимание роли контекста в дискурсивной перспективе и его ориентирующая роль в декодировании высказывания [4] и интерпретация корреляции между прагматикой и дискурсивной лингвистикой (на примере анализа театрального дискурса [5], которые вместе с тем не получили исчерпывающей интерпретации в том смысле, что не обозначены принципиальные отличия письменного от устного дискурсивного содержания, когда в устном дискурсе ведущую роль может играть интонационное решение, тогда как в письменном это преимущественно инференция и сходные механизмы. Один из таких вопросов является предметом рассмотрения в данной статье. Таким образом, отвечая на социально-лингвистический запрос, **цель** предлагаемого исследования состоит в установлении границ между познавательными установками дискурсивного анализа, с одной стороны, и прагматики – с другой, изучаются также парадигмальные корреляции в целом [1].

Поставленная цель предполагает решение следующих **задач**: выявить когнитивные основания разграничения данных направлений лингвистического анализа и предложить уточнение их границ на основе аргументации с помощью эмпирического материала.

Методика исследования основана на ряде методологических положений, убедительно обоснованных С. Т. Нефедовым и его коллегами [3; 4], которые представляются особенно значимыми для современной лингвистики. Выявляются фрагменты текста, которые можно отнести к прагматическим, основанным на семантике языковых единиц в системе языка и практически универсальных для языковых картин мира, например, высказывание *Первый спектакль был поставлен в Красноярском ТЮЗе* в силу значения инициальной фразы заставляет предположить, что после первой постановки может последовать и скорее всего последовала и вторая и н-ная (и это прагматика), тогда как фраза *Первая скрипка как всегда пришел на репетицию вовремя*, в ироническом устном дискурсе (с использованием соответствующей интонации и произнесенная в со-присутствии сообщества активных участников собрания, разделяющих знание о привычках этой первой скрипки). Таким образом отделяется прагматическая инференция от социально-культурной.

Материалом практического исследования послужили высказывания из сферы театрального дискурса (телеграм-каналы журналов «Театр» и «Театрал»), в рамках которых разграничивается прагматический и дискурсивный контент, уточнены критерии их разграничения.

При этом представляется уместным разграничивать устный и письменный формат общения: при совпадении дискурсивной информации в письменном и устном общении – например, использование пословиц, т.н. крылатых фраз различного генезиса (национальная культура, представленная в фильмах, литературе, публичных выступлениях – ср. понятное всем в свете последних событий выражение президента *Но есть нюансы*, за которым стоит многоплановое содержание), однако устный дискурс «богаче» в дискурсивном выражении, поскольку привносит информационно значимые просодические средства и актуализирует в целом информационную структуру высказывания, или его актуальное членение.

Таким образом, «именно в точке взаимодействия конвенционального значения, идущего от системы языка, с релевантными параметрами контекста происходит семантизация языковых выражений (=прагматика, уточнение наше – О. С.) и возникает актуальный, ситуативно обусловленный смысл» (=дискурсивная интерпретация – уточнение наше – О. С.) [4]. И в этом смысле можно говорить о том, что работает принцип дополнительности, когда прагматическое содержание, выводимое из лингвистического контекста [6], дополняется и финализируется дискурсивным содержанием. Их комбинирование и порождает заложенный автором смысл и позволяет интерпретировать речевое произведение. Имеет место «продолжение прагматически ориентированной лингвистики» [4, с. 85].

«Иными словами, контекст в дискурсивной перспективе выполняет (наряду с семантикой языковых структур) ориентирующую роль, служа основой дискурсивной интерпретации высказывания» [5] и позволяя спрогнозировать восприятие сообщения адресатом.

Рассмотрим, например, высказывание *В Санкт-Петербургском театре «Мастерская» расскажут о «высокой башне» и ее обитателях*, понимание которого опирается, во-первых, на отсылку к легенде о духовном восхождении, блуждающей по миру и широко известной, но и одновременно как к некоему закрытому клубу, во-вторых, для русского (в принципе хорошо подготовленного и осведомленного о литературной истории начала 20 века) читателя это также аллюзия к известной квартире Вяч. Иванова на ул. Таврической («Башне»), где бывали представители Серебряного века. В силу этого необходимо дискурсивное декодирование с опорой на национально-культурный опыт. Ср. А. Лыкова *Я люблю литературу с отсылками, которая разматывается как клубок* (29 июня 2025 – сама эта фраза автора может практически восприниматься как своего рода «манифест» дискурсивности!) – речь об игре смыслов в названии постановки К. Вагинова в Геликон-опере «Козлиная песнь»: букв. перевод с греческого слова *трагедия*. При этом автор «прислушивается к подсказкам Стравинского» в его опере по мотивам

«Похождение повесы» (1951 г.) на сюжет обличительных гравюр У. Хогарта, причем постановка состоится в зале Стравинский, что также привносит новый смысл в интерпретацию события.

Иными словами, «разматывается» комплексная игра скрытых дискурсивных смыслов, которые и позволяют адекватно расшифровать сообщение, в данном случае чрезвычайно нагруженное дискурсивным контентом и нацеленное на «своего» адресата.

Проблемы описания семантически неоднозначных контекстов и декодирования «невывраженного» на основе инференции с опорой на социокультурное знание и позволяет достичь цели коммуникации. Прагматическое знание, напротив, основано на, во-первых, эксплицитном выражении, например, отношения говорящего к описываемой ситуации и к своему речевому произведению, но также и к имплицитному знанию, которое требует лингвистически предсказуемой инференции, тогда как дискурсивный смысл декодируется с опорой на социально-культурное взаимодействие.

ЛИТЕРАТУРА

1. Cornish F. Text, discourse, context: A meta-trilogy for discourse analysis // *Journal of Pragmatics*. 2022, (199): PP. 91–104.
2. Потякайло С. А. Библиометрический контент-анализ тематики диссертационных исследований по лингвистике (2000–2023 гг.): автореф. дис. ... канд. филол. наук.
3. Нефёдов С. Т. Дискурс: направления, школы, методология дискурсивного анализа // *Немецкая филология в Санкт-Петербургском государственном университете*. 2018. № 7. С. 11–34.
4. Нефёдов С. Т., Чернявская В. Е. Контекст в лингвистическом анализе: прагматическая и дискурсивно-аналитическая перспектива // *Вестник Томского государственного университета. Сер. Филология*. 2020. № 63. С. 83–97.
5. Сулейманова О. А., Борботько Л. А. Дискурсивный анализ и прагматика: возможно ли провести границу? // *Вестник Санкт-Петербургского университета. Язык и литература*. 2024, № 21 (4). С. 919–925.
6. Ronai E., Xiang M. What could have been said? Alternatives and variability in pragmatic inferences? // *Journal of Memory and Language*. 2024. № 136. PP. 104–107.

УДК 378.8.091.313.11-029:004

Тивьяева Ирина Владимировна
доктор филологических наук,
профессор кафедры языкознания
и переводоведения
Московский городской
педагогический университет
г. Москва, Россия

Irina Tivyaeva
Advanced Doctorate in Philology,
Professor, Department of Linguistics
and Translation Studies
Moscow City University
Moscow, Russia
TivyaevaIV@mgpu.ru

ВОЗМОЖНОСТИ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ В РАЗВИТИИ ИССЛЕДОВАТЕЛЬСКОЙ КОМПЕТЕНЦИИ СТУДЕНТОВ-ПЕРЕВОДЧИКОВ

В статье исследуется потенциал больших языковых моделей GigaChat, DeepSeek и Perplexity в развитии исследовательской компетенции студентов-переводчиков. Актуальность работы обусловлена требованиями цифровизации лингвистического образования и необходимостью интеграции генеративного искусственного интеллекта в учебный процесс. Автор рассматривает функциональные возможности больших языковых моделей применительно к ключевым исследовательским задачам, включая выбор темы, составление плана работы, поиск литературы, анализ данных и оформление библиографических ссылок. Методология исследования включает тестирование моделей на основе 20 диалогов, охватывающих шесть направлений, таких как машинный перевод, постредактирование, применение искусственного интеллекта в письменном и устном переводе и др. Выводы подчеркивают целесообразность использования больших языковых моделей как вспомогательных инструментов для рутинных этапов исследования и необходимость развития методологической базы, регламентирующей применение инструментов искусственного интеллекта при подготовке лингвистов по программам бакалавриата и магистратуры.

Ключевые слова: большая языковая модель; генеративный искусственный интеллект; исследовательская компетенция; подготовка переводчиков; иноязычное образование.

THE POTENTIAL OF LARGE LANGUAGE MODELS IN DEVELOPING RESEARCH SKILLS OF TRANSLATORS-IN-TRAINING

The paper investigates the capabilities of large language models, specifically GigaChat, DeepSeek, and Perplexity, in fostering research competencies among translation students. The research is motivated by the growing digitalization of linguistic education and the imperative to incorporate generative artificial intelligence into academic training. The paper examines how these LLMs can support fundamental research activities, including topic selection, research design, literature retrieval, data analysis, and bibliographic formatting. Employing a dialog-based testing methodology (20 dialogues per model), the study evaluates LLM performance across six key domains: machine translation, post-editing, and AI applications in translation and

interpreting. The results demonstrate that while LLMs serve as effective auxiliary tools for routine research tasks, their implementation in translator education requires developing methodological frameworks for both undergraduate and graduate levels.

Key words: large language models; generative AI; research competence; translator training; foreign language education.

Одна из ключевых задач программ подготовки профессиональных лингвистов-переводчиков в университете связана с обеспечением их конкурентоспособности на рынке лингвистических услуг, который сегодня в качестве основного требования предписывает наличие у выпускника развитых умений и навыков работы в цифровой среде [1]. Адаптация к новым требованиям рынка и непрерывное профессиональное развитие являются ключевыми факторами успеха в данной сфере деятельности. В эру цифровизации квалифицированный переводчик должен быть готов не только к переводу разных типов текстов, редактированию и технической обработке материалов, но и к выполнению других рабочих задач. Современные переводчики и редакторы должны обладать широким спектром компетенций, включающим владение иностранными языками, основами межкультурной коммуникации, текстовыми технологиями, а также навыками поисковой и исследовательской деятельности [2; 3].

В соответствии с современными требованиями к подготовке лингвистов-переводчиков в вузе, в частности, в соответствии с ФГОС РФ по направлению магистратуры 45.04.02 *Лингвистика* специалист в области перевода, помимо прочего, должен обладать исследовательской компетенцией, то есть уметь работать с научной документацией, а также с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации. Важным аспектом подготовки является обучение работе с современными информационными ресурсами и технологиями. Студенты должны уметь использовать электронные библиотеки, базы данных, специализированные поисковые системы и программное обеспечение для обработки текстов. Не менее значимым также видится развитие навыков академического письма, включая написание научных статей, докладов, рецензий и аннотаций. Это способствует не только формированию исследовательской компетенции, но и подготовке студентов к успешной научной и профессиональной деятельности. Таким образом, подготовка лингвистов-переводчиков, в том числе развитие их исследовательской компетенции, требует интеграции теоретических знаний и практических навыков, что позволяет им успешно функционировать в динамично развивающейся профессиональной среде.

В связи с активным внедрением технологий генеративного искусственного интеллекта (далее генИИ) в процессы обработки текстовых данных, в том числе для целей перевода, редактирования и анализа, в академических

кругах ведутся дискуссии о трансформации иноязычного образования и необходимости регламентирования использования генИИ в учебном процессе [4; 5; 6]. В этих условиях возникает необходимость разработки методологических основ применения больших языковых моделей (далее LLM от английского *large language models*) в программах подготовки лингвистов-переводчиков в языковом вузе. Первым шагом в решении данной задачи, как видится, должен стать этап тестирования функциональных возможностей LLM с целью оценки потенциала их использования в учебном процессе. В данном случае нас будет интересовать в первую очередь возможность применения инструментов генИИ для целей развития исследовательской компетенции студентов и их навыков академического письма.

Следует отметить, что интерес к функциональным возможностям LLM применительно к решению прикладных задач в области лингвистики и лингводидактики возник относительно недавно на фоне общего роста интереса к технологиям генИИ. Исследователи и преподаватели-практики изучают различные аспекты возможного применения инструментов и средств генИИ для разработки дидактических материалов и создания учебного контента [7], в том числе в качестве симулятора носителя языка [8], для планирования, организации и оценки учебного процесса [9; 10] – подробный обзор актуальных работ в области применения технологии ГИИ при обучении иностранному языку приводится в работе [11].

В лингвистических исследованиях актуальную повестку также формируют работы, посвященные изучению возможностей больших языковых моделей применительно к различным вопросам анализа естественного языка, в том числе для обработки корпусных данных [12], постредактирования машинного перевода [13] и анализа редких языков [14]. В особое направление следует также вынести компаративные исследования, направленные на изучение естественных и сгенерированных текстов на разных языках в сравнительной перспективе [15].

В настоящей статье приводятся результаты оценки потенциала применения для указанных целей трех больших языковых моделей: GigaChat, DeepSeek и Perplexity. В фокусе нашего внимания сравнение функциональных возможностей указанных языковых моделей в статусе научного ассистента студентов-переводчиков. В ходе тестирования проводился анализ потенциала указанных языковых моделей применительно к следующим исследовательским задачам, решение которых является обязательным для студента-лингвиста в ходе подготовки выпускной квалификационной работы:

- выбор актуальной темы исследования;
- составление плана научно-исследовательской работы;
- поиск современной научной литературы по теме исследования;
- подготовка обзора научной литературы по теме исследования;

- поиск источников релевантных эмпирических данных;
- анализ эмпирических данных;
- составление аннотации и подбор ключевых слов;
- перевод метаданных на английский язык;
- формулировка названия научно-исследовательской работы;
- оформление библиографии.

Тестирование проводилось в рамках следующих исследовательских направлений, актуальных для программ подготовки профессиональных переводчиков (по 20 диалогов с каждой языковой моделью): автоматизированные программы перевода, машинный перевод, постредактирование машинного перевода, применение нейросетей в работе письменного переводчика, применение технологий искусственного интеллекта в работе устного переводчика, применение нейросетей для целей академического письма. Все диалоги велись на русском языке, за исключением заданий, связанных с переводом метаданных на английский язык и редактированием англоязычных текстов аннотаций.

В качестве промптов использовались следующие: 1. *Предложи актуальную тему исследования для дипломной работы в области [НАЗВАНИЕ НАПРАВЛЕНИЯ]. Тема работы должна соответствовать направлению подготовки 45.04.02 Лингвистика и уровню программы подготовки в магистратуре.* 2. *Составь план научно-исследовательской работы по теме [НАЗВАНИЕ ТЕМЫ] с учетом вышеназванных условий.* 3. *Подбери современную научную литературу по теме [НАЗВАНИЕ ТЕМЫ]. Список должен включать не менее 30 позиций, из них не менее 5 на иностранном языке. Предпочтительно, чтобы источники были изданы не позднее 2000 года. Предлагай только реально существующие работы.* 4. *Подготовь обзор научной литературы по теме [НАЗВАНИЕ ТЕМЫ]. Ориентируйся на условия, указанные выше.* 5. *Подбери источники языкового материала для формирования эмпирической базы исследования по теме [НАЗВАНИЕ ТЕМЫ]. Обязательно приводи ссылки на каждый источник.* 6. *Предложи алгоритм анализа эмпирических данных.* 7. *Составь аннотацию и подбери ключевые слова к научно-исследовательской работе по теме [НАЗВАНИЕ ТЕМЫ]. Используй составленный план работы.* 8. *Переведи на английский язык название работы, аннотацию и ключевые слова с учетом стилиевой нормы.* 9. *Оформи список литературы в соответствии с ГОСТ Р 7.0.100-2018.*

В таблице 1 ниже представлены оценки полученных от GigaChat, DeepSeek и Perplexity ответов на представленные выше промпты. Ответы языковых моделей, оцененные нами как корректные или адекватные, отмечены в таблице знаком «+», как некорректные / неадекватные – знаком «-», как частично корректные – знаком «+/-».

Результаты обращения к большим языковым моделям GigaChat, chatGPT 4.0 и Perplexity с промптами, соответствующими этапам работы над текстом научно-исследовательской работы

Название этапа	GigaChat	Perplexity	DeepSeek
Выбор актуальной темы исследования	+/-	+	+
Составление плана исследования	+	+	+
Подбор научной литературы	+/-	+/-	+/-
Обзор научной литературы	—	+/-	—
Поиск источников эмпирических данных	—	+/-	+/-
Составление алгоритма анализа языковых данных	+/-	+/-	+/-
Составление аннотации и ключевых слов	+	+	+
Перевод метаданных	+	+	+
Оформление библиографии по ГОСТ	+	+	+/-

В ходе сравнительного анализа результатов, продемонстрированных языковыми моделями GigaChat, Perplexity и DeepSeek, были выявлены их сильные и слабые стороны в контексте поддержки различных этапов исследовательского процесса в интересующей нас области. Результаты анализа показали, что DeepSeek и Perplexity демонстрируют стабильно высокие показатели на большинстве этапов, особенно в области структурирования и формализации научного текста. GigaChat, хотя и уступает в анализе литературы и подборе эмпирических данных, успешно справляется с организационными аспектами исследования, такими как составление плана и аннотаций.

Анализ выявил существенные ограничения LLM в ряде ключевых этапов подготовки научно-исследовательской или выпускной квалификационной работы в области лингвистики. В частности, все рассмотренные модели продемонстрировали низкую эффективность при анализе научной литературы, что свидетельствует об их ограниченной способности к критическому анализу и синтезу информации из различных источников.

Кроме того, поиск эмпирических данных также оказался сложной задачей для языковых моделей. Лишь DeepSeek и Perplexity частично справились с этой задачей, в то время как GigaChat показал крайне низкую

результативность. Данный результат возможно объяснить тем фактом, что большие языковые модели оперируют преимущественно готовыми текстами, не имея доступа к реальным данным, таким как корпуса языковых примеров или базы эмпирических исследований.

Несмотря на выявленные ограничения, инструменты генИИ обладают рядом существенных преимуществ, особенно в решении формальных задач научного исследования. Оформление библиографии, перевод метаданных и составление плана исследования выполняются всеми рассмотренными моделями на достаточно высоком уровне благодаря четким алгоритмам и стандартизации данных. Особого внимания заслуживают гибридные инструменты, в нашем случае это диалоговая система Perplexity, которая сочетает функции поиска и генерации текста. Perplexity демонстрирует более высокую эффективность в подборе литературы и источников данных по сравнению с другими рассмотренными решениями.

Представленный выше краткий обзор возможностей генИИ позволяет прийти к заключению, что при развитии исследовательской компетенции студентов-переводчиков в эпоху цифровизации игнорировать потенциал больших языковых моделей не представляется возможным. В отсутствие инициативы в отношении применения технологий генИИ со стороны образовательной организации обучающиеся будут предпринимать самостоятельные попытки применять генИИ в исследовательской работе, при этом без руководства со стороны наставников налицо все риски их некорректного использования в качестве генераторов готового текста без должного критического анализа полученных результатов.

Современные инструменты и технологии предоставляют будущим специалистам в области перевода широкие возможности для подготовки к разнообразным вызовам, ожидающим их в профессиональной деятельности. В условиях стремительного развития технологической среды и внедрения новейших цифровых инструментов, включая средства генИИ, современный переводчик должен обладать высокой степенью адаптивности и релевантными цифровыми навыками. В этом контексте формирование исследовательской компетенции на основе интеграции цифровых ресурсов и возможностей искусственного интеллекта становится неотъемлемой частью профессиональной подготовки переводчиков в высших учебных заведениях. На данном этапе внедрения технологий генИИ в систему иноязычного образования важным видится проектирование методологической базы, регламентирующей применение генИИ для подготовки бакалавров и магистров в языковом вузе.

ЛИТЕРАТУРА

1. Гулиянц А. Б., Гулиянц С. Б., Иванова А. М. Применение больших лингвистических данных в дидактике и профессиональной подготовке

- переводчика // Профессионально ориентированный перевод: реальность и перспективы : Сборник научных трудов, Москва, 08–10 апреля 2024 года. Москва: Российский университет дружбы народов (РУДН), 2024. С. 71–88.
2. Иванова А. М. Стратегии предпереводческого анализа специализированных текстов, подлежащих переводу с применением систем автоматизированного перевода // Язык и культура в глобальном мире : сборник статей. Санкт-Петербург : ООО "Издательство "Лема", 2025. С. 258–265.
3. Сулейманова О. А., Карданова-Бирюкова К. С., Беклемешева Н. Н. Уровневая концепция формирования уверенного переводчика в высшей школе // *Didactica Translatorica*. 2022. № 1. С. 20–25.
4. Тарева Е. Г., Тивьяева И. В. Трансформация лингводидактики: как искусственный интеллект формирует лингвообразовательные приоритеты // Современная {цифровая} дидактика : Монография. Москва : «Издательство Интеллект-Центр», 2024. С. 37–44.
5. Terra autonoma: предопределяя будущее иноязычного образования в автономном вузе : Коллективная монография / Е. Я. Орехова, Е. Г. Тарева, С. В. Михайлова [и др.] ; Под общей редакцией Е. Я. Ореховой, Е. Г. Таревой, С. В. Михайловой. Москва : Общество с ограниченной ответственностью «Языки Народов Мира», 2022. 167 с.
6. Тивьяева И. В., Михайлова С. В., Казанцева А. А. Регламентирование использования средств генеративного искусственного интеллекта в выпускной квалификационной работе // Вестник МГПУ. Серия: Филология. Теория языка. Языковое образование. 2024. № 2(54). С. 202–218. DOI 10.25688/2076-913X.2024.54.2.15.
7. Авраменко А. П., Тишина М. А. Дидактический потенциал лингвистических корпусов на базе технологий искусственного интеллекта для адаптации учебных материалов // Вестник Московского государственного областного университета. Серия: Педагогика. 2023. № 1. С. 29-38. DOI 10.18384/2310-7219-2023-1-29-38.
8. Сильчева А. Г., Ламзина А. В., Павлова Т. Л. Особенности использования текстовых и графических чат-ботов с искусственным интеллектом в преподавании английского языка // Перспективы науки и образования. 2023. № 4(64). С. 621–635. DOI 10.32744/pse.2023.4.38.
9. Вишневецкая Н. В. Искусственный интеллект на уроке иностранного языка // Три «Л» в парадигме современного гуманитарного знания: лингвистика, литературоведение, лингводидактика : Сборник научных статей Всероссийской научно-практической конференции, Москва, 23 ноября 2023 года. Москва: ООО «Языки Народов Мира», 2024. С. 364–369.
10. Вишневецкая Н. В., Фетисова А. А. Синтез мобильных технологий и искусственного интеллекта как основа развития коммуникативной культуры будущего преподавателя иностранного языка // Диалог культур-культура диалога в многонациональном городском пространстве : Материалы Четвертой международной научно-практической конференции, Москва, 27 февраля 01 2024 года. Москва: ООО «Языки народов мира», 2024. С. 90–98.

11. Сысоев П. В. Использование технологий искусственного интеллекта в обучении иностранному языку: тематика методических работ за 2023 год и перспективы дальнейших исследований // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2024. Т. 29, № 2. С. 294–308. DOI 10.20310/1810-0201-2024-29-2-294-308.
12. Пронина Е. В., Пронин Д. Д. Исследовательский потенциал изучения корпуса произведений русской литературы с помощью цифровых лингвистических методов и технологий искусственного интеллекта (проект Lensky) // Современный ученый. 2023. № 3. С. 92–105.
13. Moneus A. M., Sahari Y. Artificial intelligence and human translation: A contrastive study based on legal texts [Electronic resource] // Heliyon. 2024. Vol. 10, Issue 6, e28106. DOI: <https://doi.org/10.1016/j.heliyon.2024.e28106>.
14. Новикова М. Л., Новиков Ф. Н. Анализ возможностей использования алгоритмов больших языковых моделей применительно к языкам народов РФ и СНГ в современном цифровом пространстве // Филологические науки. Научные доклады высшей школы. 2023. № S6. С. 3–11.
15. Тельпов Р. Е., Ларцина С. В. Типовые различия естественных и сгенерированных нейронной сетью текстов в квантитативном аспекте // Научный диалог. 2023. Т. 12, № 7. С. 47–65.

УДК 81 '27

Ширяева Татьяна Александровна
доктор филологических наук,
заведующая кафедрой английского
языка и профессиональной
коммуникации
ФГБОУ ВО «Пятигорский
государственный университет»
г. Пятигорск, Россия

Tatyana Shiryayeva
Doctor of Philological Sciences,
Head of the Department of English
and Professional Communication
Pyatigorsk State University
Pyatigorsk, Russia
shiryayevat@list.ru

КОРПУСНЫЙ АНАЛИЗ ЦЕННОСТНЫХ КОНЦЕПТОВ В АНГЛОЯЗЫЧНОМ IT-ДИСКУРСЕ

Статья представляет всесторонний корпусный анализ аксиологических концептов в англоязычном IT-дискурсе, выявляя их лексические и прагматические особенности. На основе методов корпусного анализа, лингвистической семантики и прагматики исследуются структурные уровни и социальной значимости ключевых ценностных единиц, таких как DATA, PRIVACY и SECURITY, а также других концептов, формирующих системы ценностных ориентиров профессионального сообщества и широкой аудитории. Особое внимание уделяется анализу лексем с прагма-аксиологической нагрузкой, включая нейтральную и эмоционально окрашенную лексику, а также их функциональной роли в формировании общественного сознания. Полученные результаты демонстрируют многоуровневую организацию ценностных концептов и их активное участие в коммуникативных практиках современного цифрового общества, что подчеркивает значимость дискурсивных исследований в понимании трансформации социокультурных ориентиров в эпоху цифровых технологий.

Ключевые слова: концепт; аксиология; ценности; дискурс; лексика; цифровое общество; влияние.

CORPUS-BASED ANALYSIS OF VALUE CONCEPTS IN THE ENGLISH IT-DISOURSE

This article presents a comprehensive corpus-based analysis of axiological concepts in the English-speaking IT discourse, highlighting their lexical and pragmatic features. Utilizing corpus analysis methods, linguistic semantics, and pragmatics, the study examines the structural levels and social significance of key value units such as DATA, PRIVACY, and SECURITY, as well as other concepts that shape the system of values and orientations within the professional community and the broader audience. Particular attention is given to the analysis of lexemes with praxeological-axiological loads, including neutral and emotionally connotated vocabulary, along with their functional roles in shaping public consciousness. The findings demonstrate the multilayered organization of value concepts and their active participation in communicative practices of modern digital society, emphasizing the importance of discursive research in understanding the transformation of sociocultural guidelines in the era of digital technologies.

Key words: concept; axiology; values; discourse; lexicon; digital society; influence.

В условиях стремительных технологических изменений и постоянной цифровой революции, фундаментально преобразующих структуру общества и сложившиеся формы коммуникации, особое значение приобретает иссле-

дование IT-дискурса как динамичного и многоаспектного явления, оказывающего значительное влияние на практически все сферы жизнедеятельности современного социума. Сегодня как отечественные, так и зарубежные лингвисты всё чаще обращаются к анализу дискурса в контексте его роли в формировании и отражении социокультурных изменений, связанных с цифровой эпохой. Не вызывает сомнений, что в эпоху цифровых трансформаций язык по-прежнему служит не только средством коммуникации, но и мощным инструментом формирования общественного сознания, идентичности и системы ценностей. Особое значение в этом контексте приобретает анализ IT-дискурса, активно проникающего в различные сферы жизнедеятельности [1] и являющегося важной областью лингвистических исследований. Одной из особо острых проблем, находящихся в центре внимания современных ученых, является выявление и интерпретация ценностных концептов – аксиологических единиц, выражающих важнейшие социальные и культурные ориентиры [2].

Актуальность исследования ценностных концептов в англоязычном IT-дискурсе обусловлена тем, что в условиях цифровых изменений данный тип дискурса становится одним из наиболее значительных носителей не только технологической информации, но и культурных ценностей, особенностей общественного восприятия и идеологических установок. В рамках данной работы основное внимание сосредоточено на вскрытии и анализе ценностных концептов, которые выражаются через лексические средства с прагма-аксиологической окраской, формируя систему аксиологических ориентиров специалистов и широкой аудитории. Конечная цель заключается в анализе структурных и прагматических свойств данных концептов с применением корпусных методов, что позволяет выявить их типологию, многослойность и социальную значимость.

Достижение поставленной цели предполагает опору на комплексное методологическое основание, объединяющее корпусный анализ, лингвистическую семантику и прагматику, а также социокультурные подходы. Исследование базируется на корпусе англоязычного IT-дискурса, включающего научно-популярные статьи, технические описания, профессиональный сленг и дискурсивные практики, связанные с информационными технологиями. В качестве существенных ориентиров в работе использованы концепции, выявленные в предыдущих работах, проведенных в рамках анализа IT-дискурса [3, 4]. Особое внимание уделено лексическим средствам с прагма-аксиологическим зарядом, охватывающим нейтральную лексику (“data set”, “data analysis”, “data storage”) и эмоционально окрашенным, разговорным или сленговым выражениям, для выявления которых применялись автоматизированные инструменты корпусных исследований, такие как критерии конкорданса, кластерный анализ, частотный анализ и контекстуальные исследования, что позволило систематизировать употребление данной лексики и описать контекстуальные особенности функционирования концептов. Общая

модель исследования предполагает описание не только лексического наполнения концептов, но и их когнитивных, прагматических и социально-культурных аспектов.

В рамках проведенного анализа было выявлено 12 ключевых аксиологических концептов, среди которых наиболее часто встречаются DATA, PRIVACY и SECURITY, являющиеся ядром системы ценностных ориентаций в англоязычном IT-дискурсе. Наряду с ними выделяются такие концепты, как OPENNESS, TRANSPARENCY, INNOVATION, INTEGRITY, TRUST и RESPONSIBILITY, каждый из которых воплощается через соответствующие лексические выражения и играет роль в структурировании профессиональной и общественной дискурсивных практик.

Структура концептов, выявленных в корпусе, характеризуется многоуровневой организацией. На когнитивном уровне концепты выступают как образные единицы, отражающие фундаментальные ценностные ориентации, такие как конфиденциальность, безопасность, инновации и доверие. Лексическая реализация этих концептов включает как нейтральную терминологию (“data security”, “privacy policy”), так и эмотивные выражения и сленг (“data breach”, “privacy concerns”, “cybersecurity threats”). В свою очередь, в прагматической плоскости данные концепты функционируют как инструменты интерпретации и оценки технологических явлений, влияя на восприятие аудитории, укрепляя или переопределяя ценностные установки специалистов и пользователей.

Для анализа использовался специализированный корпус объемом более 2 миллионов слов, включающий публикации авторитетных изданий, блоги, техническую документацию и научно-популярные материалы последних десяти лет. В процессе корпусного анализа были выявлены частотные характеристики лексем, их распространенность в различных типах дискурса, а также контекстуальные связи. Так, для каждого ключевого концепта “PRIVACY” («конфиденциальность»), осуществлялся анализ степени его употребления, сфер применения и ассоциативных связей. В частности, слово “privacy” демонстрирует широкий спектр употреблений, включающих формализованные выражения типа “privacy settings” (настройки конфиденциальности), “data privacy” (конфиденциальность данных), а также эмоционально насыщенные выражения, такие как “privacy nightmare” (кошмар для конфиденциальности) и “privacy invasion” (посягательство на личную жизнь). Такой языковой спектр свидетельствует о многогранности функционирования концепта: с одной стороны, он структурирует ценностные ориентации в области информационной безопасности; с другой – формирует эмоциональность и реактивность аудитории, что указывает на важность концепта в формировании общественного мнения и культурных моделей.

Аналогичным образом концепт “SECURITY” подвергался детальному исследованию, демонстрируя активную вербализацию в профессиональных терминах, таких как “cybersecurity” (кибербезопасность), “security protocols”

(протоколы безопасности), а также в сленговых выражениях, например, “security breach” (утечка безопасности), “security glitch” (сбой безопасности). Анализ показал, что такие лексемы не только передают прагматическую функцию – описывать угрозы и меры защиты, – но и насыщены эмоциональной окраской, что указывает на их роль в формировании социального восприятия рисков и тревог, связанных с кибербезопасностью.

Использование корпусных методов позволило выявить не только частотные и распределительные свойства лексем, но и установить ассоциативные связи, отношения с соседними по контексту единицами и выявить типичные коллокации, что, безусловно, способствует глубокому пониманию динамики функционирования ценностных концептов в современном IT-дискурсе [5]. Такой подход способствует не только количественному описанию лингвистических характеристик, но и качественному осмыслению их семантической и прагматической нагрузки, что является критически важным для построения целостных моделей коммуникационных стратегий в сфере информационных технологий.

Не менее интересным итогом работы явился тот факт, что проведенный анализ позволил не только идентифицировать ключевые ценностные концепты в англоязычном IT-дискурсе, но и выявить их комплексную дискурсивную функцию, что существенно расширяет понимание роли лингвистических элементов в формировании социального и культурного пространства цифровизации. Анализ показал, что данные концепты выполняют три важнейшие функции в структуре профессиональной коммуникации: во-первых, структурируют профессиональный язык, делая его более осмысленным, систематизированным и соответствующим современным вызовам и реалиям; во-вторых, формируют эмоциональный фон взаимодействия, предопределяя тем самым характер интеракций и уровень доверия между участниками дискурса; и, в-третьих, участвуют в конструировании и репрезентации социального и культурного контекста, способствуя закреплению коллективных ценностных ориентиров и образцов поведения.

Рассмотрим более подробно, как эти функции реализуются на практике. Например, концепт “TRANSPARENCY” часто используется в контекстах, связанных с этическими аспектами внедрения новых технологий, в том числе в дискурсах об искусственном интеллекте, блокчейне и других инновационных направлениях. Его актуализация в подобных контекстах способствует формированию позитивного образа технологий, ассоциирующегося с открытостью, честностью и ответственностью, что в свою очередь усиливает доверие пользователей и общества в целом к новым цифровым решениям. Не вызывает сомнений, что в контексте IT-дискурса концепт “TRANSPARENCY” выступает не только как лингвистическая единица, но и как дискурсивный инструмент, закрепляющий ценностные нормы в профессиональной сфере.

Совершенно иная ситуация складывается с концептом “PRIVACY”, который зачастую используется в связи с острыми спорами и публичными дискуссиями, связанными с ростом числа случаев утечек данных, злоупотреблений и усилением государственного и корпоративного наблюдения. Вербализация данного концепта в изучаемом дискурсе вызывает противоположные реакции: с одной стороны, требования защиты личной информации, с другой – необходимость баланса между безопасностью и свободой, что способствует формированию конкретных ценностных приоритетов и этических стандартов в обществе. Именно через актуализацию данного концепта формируется коллективный образ цифровизации, отражающий понимание социальной ответственности, жизненно важную для формирования доверия, а также механизмы регулирования и ответственности в цифровую эпоху.

Следует подчеркнуть, что использование данных ценностных концептов способствует не только формированию индивидуальных и коллективных представлений о технологии, но и активному участию в создании общего понимания современной жизнедеятельности, охватывая такие важные социальные сферы, как этика, безопасность, инновации и доверие. В итоге, данный дискурс не просто отражает существующие ценности, но и активно их конструирует, воздействуя на формирование коллективного сознания и восприятия технологического прогресса в широком социальном контексте.

Итак, подводя итоги, подчеркнем, что проведенный корпусной анализ ценностных концептов в англоязычном IT-дискурсе показал, что данные единицы выполняют важные когнитивные, прагматические и социальные функции, выступая в качестве активных инструментов формирования общественного сознания. Они объединяют лингвистические средства с идеологическими и культурными ценностями, создавая систему ценностных ориентиров, которая актуализируется и трансформируется в условиях постоянных технологических новшеств. Использование корпусных методов позволило не только систематизировать и глубже понять механизмы функционирования этих концептов, но и определить их роль в современном социуме.

ЛИТЕРАТУРА

1. Baker, Paul. 2018. Corpora and the study of evaluative language in digital discourse. *Journal of Pragmatics*, 134, 124–137. DOI: 10.1016/j.pragma.2018.07.015.
2. Шустрова, Елена Валерьевна. 2020. Когнитивные аспекты оценочной семантики терминологических единиц в информационно-коммуникационных технологиях. *Вестник Томского государственного педагогического университета*, № 4 (212), 106–112. DOI: 10.23951/1607-6156-2020-4-106-112.
3. Čermák, Patrik & Pavel Švec. 2020. Analyzing evaluative concepts in English IT discourse: A corpus approach. *International Journal of Corpus Linguistics*, 25(2), 255–277. DOI: 10.1075/ijcl.00041.cep.

4. Лебедева, Светлана Вячеславовна. 2021. Корпусной анализ ценностных концептов в англоязычном информационном IT-дискурсе. Вестник МГГУ. Серия «Лингвистика», № 4, 124–138. DOI: 10.22206/vestnik.2021.4.124.
5. Кузнецова, Марина Алексеевна. 2020. Лексические средства выражения ценностей в английском IT-дискурсе: корпусный анализ. Вестник Московского государственного университета. Серия 19. Лингвистика и международные коммуникации, № 6, 45–59. DOI: 10.18688/2079-7892-2020-6-45–59.

Секция 1. КОРПУСНЫЕ ИССЛЕДОВАНИЯ ЯЗЫКА И РЕЧИ

УДК 811.41

Алексеева Мария Евгеньевна
кандидат филологических наук,
старший преподаватель
СПбГУ
г. Санкт-Петербург, Россия

Maria Alekseeva
PhD in Philology,
Senior lecturer
St. Petersburg State University
St Petersburg, Russia
m.e.alekseeva@spbu.ru

ЗНАЧЕНИЕ ПРАГМАТИЧЕСКОГО КОНТЕКСТА И РЕФЕРЕНЦИАЛЬНОГО СТАТУСА В ОФОРМЛЕНИИ ПРЯМОГО ДОПОЛНЕНИЯ С КВАНТИФИКАТОРОМ *KOL*: КОРПУСНЫЙ АНАЛИЗ ДАННЫХ СОВРЕМЕННОГО ИВРИТА

Статья посвящена классификации конструкций с квантификатором *kol* в современном иврите и вариативности их объектного маркирования с учетом референциального статуса и прагматических факторов. Исследование показывает, что *kol* включено в состав различных по статусу референциальных выражений, от нереферентных до конкретно-референтных. Контекст употребления влияет на интерпретацию таких конструкций не менее значительно, чем формальные элементы структуры: в зависимости от интерпретации *kol* может быть понято как 'весь, все, всё, каждый, любой'. Используя методы и инструменты корпусной лингвистики, автор демонстрирует, какие факторы лицензируют объектное маркирование конструкций с квантификатором *kol*, проверяя гипотезу о том, что наличие, в частности, определенного артикля, не является определяющим фактором в этом вопросе. В пример приводится вариативность падежного оформления *kol/ha-kol*, где лицензирующим маркирование фактором, по-видимому, является именно референциальный статус. На данных корпуса показана также существенная вариативность объектного маркирования конструкций *kol mi še-* и *kol ta še-* (100 % и 55 %, соответственно) и прагматические факторы, одушевленность референта в данных примерах играют ключевую роль. Результаты исследования предварительно подтверждают гипотезу о значимости референциальных характеристик и прагматического компонента как факторов-лицензоров объектного маркирования дополнений с квантификатором *kol* в современном иврите и указывают на необходимость дальнейших исследований этого явления.

Ключевые слова: иврит; семитские языки; квантификатор; объект; референция; прагматика; одушевленность.

THE ROLE OF PRAGMATIC CONTEXT AND REFERENTIAL STATUS IN MARKING DIRECT OBJECTS WITH THE QUANTIFIER *KOL*: CORPUS ANALYSIS OF MODERN HEBREW DATA

The article focuses on the classification of constructions with the quantifier *kol* in Modern Hebrew and their differential object marking, accounting for the referential status and pragmatic factors. The study shows that *kol* is included in the composition of referential expressions of various statuses, from non-referential to specific. The context influences the interpretation of such constructions no less significantly than the formal elements: depending on the

interpretation, *kol* can be understood as 'all, every, each, any'. Using the methods and tools of corpus linguistics, the author demonstrates which factors license the object marking of constructions with the quantifier *kol*, testing the hypothesis that the use of a definite article, in particular, is not a determining factor in this matter. An example is the variability of the case marking of *kol/ha-kol*, where the licensing factor for marking seems to be precisely the referential status. The corpus data also shows significant variability in the object marking of the constructions *kol mi še-* and *kol ma še-* (100 % and 55 %, respectively) and pragmatic factors, the animacy of the referent play a key role in these cases. The results of this preliminary study confirm the hypothesis about the significance of referential characteristics and the pragmatic component as licensing factors for the object marking of objects coded with the quantifier *kol* in Modern Hebrew and indicate the need for further research into this phenomenon.

Key words: Hebrew; Semitic languages; quantifier; object; referentiality; pragmatics; animacy.

Несмотря на отсутствие подробной классификации способов оформления референтных и нереферентных именных групп (ИГ) в иврите и в целом недостаточную изученность темы референциальности на материале современного иврита, некоторые исследователи используют теоретические разработки данной области лингвистики в своих исследованиях. Но большинство авторов упоминают о референтности или любых характеристиках, связанных с прагматическим контекстом, вскользь, руководствуясь, прежде всего, формальными показателями.

В данной статье объектом исследования являются конструкции с квантификатором *kol* в современном иврите, традиционно называемые универсальными. Универсальные ИГ могут быть представлены как конкретными, так и абстрактными множествами. Исследование ставит целью определить, зависит ли в современном иврите способ объектного маркирования конструкций с квантификатором *kol* от прагматического контекста высказывания и референциального статуса данной ИГ. Выполнено данное исследование на базе корпуса современного иврита Hebrew Web 2021 (heTenTen21), расположенного на платформе SketchEngine (sketchengine.eu).

В современном иврите для кодирования как конкретных, так и абстрактных множеств частотно используется квантификатор *kol* 'все', сопровождаемый формой имени во множественном числе: *kol (ha-)sfarim* 'все книги'. В работе известного лингвиста Л. Глинерта отмечается, что универсальная интерпретация для ИГ, включающих квантификатор *kol*, возможна как в случае, если существительное оформлено показателем определенности, т. е. если «объемлющее» множество состоит из нескольких объектов, введенных в поле зрения адресата речи или однозначно определяемых выбранной дескрипцией, так и в том случае, когда имя не оформлено показателем [1, р. 72]. Однако в последние годы, как показывают данные корпуса, все более нормативным является употребление квантификатора *kol* с формой множественного числа в определенном состоянии.

Отметим, что на корреляцию параметров «определенность» и «референциальный статус» указывали многие исследователи [2; 3; 4], основываясь, в том числе, на материале структурно различных языков [5; 6].

Квантификатор *kol* может употребляться не только для обозначения «объемлющего» множества. При употреблении с именем существительным в форме единственного числа, кодирующим референт с конкретно определенной референцией и оформленным, соответственно, показателем определенности (например, артиклем), вся ИГ также приобретает конкретно-референтный статус и значение полноты охвата *kol ha-sefer* ‘вся книга’. При необходимости подчеркнуть полный охват объекта действием в постпозиции к имени используется детерминатив *kulo* (ms) (*kula* (fs)) ‘целиком’.

Наличие/отсутствие показателя определенности (определенного артикля) у существительного в ИГ с квантификатором *kol* будет играть ключевую роль для интерпретации всей ИГ. Употребление *kol* с неопределенным именем существительным в форме единственного числа требует индивидуального рассмотрения элементов множества, участвующих в описываемой ситуации, т. е., в терминологии А. Д. Шмелева выражает значение «выделенности» [7, с. 87], соответствующее русскому слову «каждый» (*kol sefer* ‘каждая/любая книга’).

Причем в данном случае множество, из которого выделен объект речи, может быть как известным адресату и конечным (рус. «каждый»), так и бесконечным, подразумевая всё множество объектов, называемых данных именем (близко к семантике рус. «всякий/любой»). В отличие от русского языка, различие значений «каждый» (индивидуально рассматриваемый элемент открытого или ограниченного множества) и «всякий» (собирательное понимание всех элементов принципиально неограниченного множества) на лексическом уровне в иврите не производится, в любом из приведенных выше примеров правомерным будет употребление словосочетания *kol sefer*. Однако выявить точное значение такой ИГ и, соответственно, ее референциальный статус (дистрибутивная множественность – конкретно-референтный, а открытый экстенционал – универсальный [8, с. 96]) позволяет контекст.

(1)	<i>im</i>	<i>tiš'alu</i>	<i>kol</i>	<i>xatan</i>	<i>bar micva</i>
	если	спрашивать.FUT.2PL	весь	жених	бар-мицва
	‘Если вы спросите любого празднующего бар-мицву ,				
	<i>ta</i>	<i>ha-davar</i>	<i>še-hu</i>	<i>haxi</i>	<i>mexake</i>
	что	DEF-вещь	который-он	наиболее	ждать.PRS.MSG
	чего он больше всего ждет,				
	<i>hu</i>	<i>yagid</i>	<i>la-xem</i>	<i>matanot</i>	
	он	сказать.FUT.3MSG	DAT-2MPL	подарки	
	он скажет вам – подарки’.				

В данном примере (1) единственной возможной интерпретацией выражения *kol xatan bar micva* является именно ‘любой празднующий бар-мицву’, поскольку индивидуальное рассмотрение каждого объекта (в данном случае, опрос каждого празднующего бар-мицву мальчика) невозможно, да и не предполагается контекстом высказывания. Соответственно, ИГ *kol xatan bar micva* мы будем воспринимать как ИГ с универсальной референ-

цией, а анафорические местоимения *hu* (ms) – как имеющую *переменную* референцию в релевантном-денотативном пространстве с обобщенным значением.

А. Д. Шмелев, однако, также связывает такой тип высказываний с «переменным денотативным пространством», которое «задается поочередно выбираемыми элементами некоторого введенного в рассмотрение множества» [7, с. 98]. Соответственно, Шмелев вводит понятие «*дистрибутивной определенности*» для случаев, когда референт будет определяться однозначно в каждом отдельно взятом денотативном пространстве (например, «*Конь иногда сбивает седока*»), и «*дистрибутивной неопределенности*», для случаев, когда в каждом из меняющихся денотативных пространств имеет место неопределенная референция (например, «*У любых двух людей найдется тема для разговора*») [7, с. 98–99].

Например, дистрибутивная определенность имеет место в высказываниях, где уточнение референта происходит с помощью притяжательного суффикса, отсылающего к упомянутому в претексте множеству:

(2)	batey	ha-din	ha-rabanim	ka-cafui
	дом.GEN	DEF-закон	DEF-раввинские	как-ожидаемо
	‘раввинские суды, как и ожидалось,			
	yaf‘ilu	’et	kol	kox-am
	делать.FUT.3PL	ACC	все	сила-POSS.3MPL
	приложат все свои силы’			

В данном случае ИГ *batey ha-din ha-rabbaniim* ‘раввинские суды’ имеет конкретно-референтный статус в переменном денотативном пространстве, поскольку подразумевается, что *каждый* конкретный суд (даже если он не упоминался ранее в претексте и не адресат не может все их однозначно идентифицировать) приложит все *свои* силы, чтобы не допустить описанной в претексте ситуации.

В примерах (1–2) прослеживается корреляция между оформлением ИГ в позиции прямого объекта маркером аккумулятива *’et* и референциальным статусом ИГ, не сводимому при анализе только к наличию/отсутствию формальных показателей, а требующим понимания контекста высказывания и места соответствующего референта в релевантно-денотативном пространстве.

Важнейшим фактором при определении референциального статуса ИГ вообще и, в том числе, ИГ с *kol* является употребления ИГ в контексте отрицания. Причем в случае, если отрицание заключено в семантике глагола, то *kol* чаще всего формирует ИГ универсальной референции (3), а если отрицается само действие, выражаемое глаголом, то ИГ будет считаться нереферентной (4).

(3)	<i>haras</i>		<i>kol</i>	<i>efšarut</i>	
	разрушать.PST.3MSG		весь	возможность	
	‘Он разрушил любую возможность.’				
(4)	<i>hu</i>	<i>lo</i>	<i>kibel</i>	<i>kol</i>	<i>tašlum</i>
	он	не	получать.PST.3MSG	весь	плата
	‘Он не получил никакой платы.’				

Рассмотри также вариативность объектного маркирования местоимения всеобщности *kol*, кодирующего референт самостоятельно и соответствующего по значению местоимению «всё» в русском языке.

В соответствии с грамматикой Л. Глинерта, в качестве прямого дополнения могут употребляться две формы местоимения: с определенным артиклем (*ha-kol*) и без него (*kol*), но в обоих случаях такой объект не маркируется [1, р. 94]:

(5)	<i>ra'iti</i>	<i>(ha-)kol.</i>
	видеть.PST.1SG	(DEF-)весь
	‘Я видел(а) всё.’	

В целом, употребление двух форм для указания на бесконечное множество («всё»), а также практически полный запрет на маркирование объекта, кодируемого таким образом, подтверждается данными исследовательского корпуса. Корпус демонстрирует также тенденцию регулярного употребления формы *kol*, как вершины придаточного предложения, формально ограничивающего бесконечное множество:

(6)	<i>e'ese</i>	<i>kol</i>	<i>še-</i>	<i>be-</i>	<i>yad-ai</i>
	делать.FUT.1SG		весь	что	в- руки-POSS.1SG
	‘Я сделаю все, что в моих руках.’				
(7)	<i>nisiti</i>	<i>kol</i>	<i>še-</i>	<i>be-</i>	<i>yexolt-i</i>
	пытаться.PST.1SG	весь	что	в-	возможность.GEN-POSS.1SG
	‘Я пробовал (сделать) все, что в моих силах.’				

Похожее значение относительных придаточных наблюдается во всех зафиксированных контекстах употребления *kol*. Местоимение *ha-kol* также частотно кодирует Пациенс глагола ‘*asa* ‘делать’, однако в большинстве случаев (9/10) не требует присоединения придаточного предложения, как-либо ограничивающего указанное множество. Вместо относительного придаточного за местоимением следует формулировка цели того действия, которое описывает финитная форма глагола, вводящаяся союзом *kdey* ‘чтобы’, а также обстоятельство действия или его бенефициант, кодирующийся с помощью предложной группы. Как ни странно, но регистр речи также, по-видимому, не оказывает существенного влияния, на выбор способа кодирования референта: *kol* – 4 в формальном регистре, 3 – в неформальном, *ha-kol* – 6 в формальном регистре, 4 – в неформальном. Таким образом, представляется, что ни одна из перечисленных характеристик не оказывают влияния на способ объектного маркирования данного местоимения всеобщности.

Из 17 контекстов употребления *kol/ha-kol* маркируется только один (6 %), в котором употребляется *ha-kol*. С одной стороны, наличие определенного артикля может показаться достаточным основанием для маркирования (как указано в грамматиках иврита относительно любых ИГ), однако в остальных 9 контекстах *ha-kol* не маркируется, несмотря на наличие показателя. Мы видим объяснение такой вариативности в различном референциальном статусе местоимений в этих контекстах. В 16 контекстах статус

местоимений *kol/ha-kol* является универсальным, и только в 1 – конкретно-референтным. Именно этот единственный контекст демонстрирует наличие маркера аккузатива *'et* (8).

В описании рецепта приготовления салата употребляется высказывание:

- (8) *kocecim* *'et* *he-'alim* *u-me'arbevīm* *'et* *ha-kol*
 резать.PRS. ACC DEF-листья и-смешивать.PRS.MPL ACC DEF-всё
 MPL
 'Режем (эти) листья и смешиваем (это) всё.'

Местоимение *ha-kol* в (8) является анафорическим и указывает на перечисленные ранее в претексте ингредиенты. Соответственно, оно обладает конкретно-референтным статусом, который, на наш взгляд, в данном примере и лицензирует объектное маркирование.

Рассмотрим выражение *kol 'exad me-* 'каждый из...'. В контексте (9) данная ИГ будет интерпретироваться как конкретно-референтная, т.к. все референты указанного множества известны из претекста и действие транзитивного глагола будет распространяться на каждый элемент множества:

- (9) *lenaseax* *'et* *ktav* *ha-išum*
 формулировать.INF ACC документ DEF-обвинение
 '(чтобы) сформулировать (этот) обвинительный акт
hi bikša *lefaked* *'et* *kol* *'exad*
 она просить.PST.3FSG приказывать.INF ACC весь один
 она просила предъявить (обвинение) **каждому**
me-arba'at *ha-ne'ešamim* *al-xelk-o* *bilvad* *ba-paraša*
 из-четыре.GEN DEF-обвиняемые на-часть- только в-DEF-дело
 POSS.3MSG
 из четырех подсудимых только по его части в (этом) деле'

Сравним данный контекст (9), где референциальное выражение указывает на индивидуализированную группу объектов, упомянутых в предшествующем дискурсе, в которой каждый из объектов будет претерпевать действие, обозначаемое глаголом транзитивной клаузы, и нижеследующий гипотетический контекст (10), в котором референциальное выражение *kol 'exad me-* имеет универсальное значение и не маркируется аккузативом.

- (10) *kdey* *še-kol* *'exad* *me-hem* *yevi*
 чтобы что-весь один из-них приводить.FUT.3MSG
 'для того, чтобы каждый из них привел
rešimat *šemot* *"manhigey* *ha-mexa'a"*
 список.GEN имена лидер.MPL.GEN DEF-протест
 список имен «лидеров протеста»
ve-ye'afyen *kol* *'exad* *me-hem*
 и-характеризовать.FUT.3MSG все один из-них.MPL
 и охарактеризовал **каждого из них...**'

В примере (10) говорящим не только используется контекст снятой утвердительности (сослагательное наклонение и гипотетическая ситуация, которую говорящий считает маловероятной, что очевидно из контекста высказывания), но и множество, к которому относится выражение «каждого

из них» не является конкретно-референтным, несмотря на употребление определенного артикля (*manhigey ha-teha 'a* ‘лидеры DEF-протеста’). Наличие артикля не обязательно предполагает конкретно-референтный статус ИГ, а в данном случае автор сам применяет дополнительное средство, указывающее не неопределенность референтов множества ‘лидеры протеста’, давая возможность трактовать данное множество как открытый экстенсионал, – заключает данную ИГ в кавычки. И, таким образом, структурно одинаковые ИГ в примерах (9) и (10) маркируются принципиально различно.

Однако корпус демонстрирует еще большую вариативность объектного маркирования ИГ с квантификатором *kol*, в которых формальной вершиной относительного придаточного становится относительное местоимение *mi* или *ta*, кодируя референт конструкцией типа «всё то, что...» / «все те, кто...».

Тип референциального выражения	Маркировано	%
<i>kol mi še-</i> ‘каждый/все кто ...’	4/4	100 %
<i>kol ta še-</i> ‘всё что ...’	11/20	55 %
Общий итог	15/24	63 %

Из 4 маркированных референциальных выражений, включающих местоимение *mi* ‘кто’, 3 обладают конкретно-референтным статусом, и 1 – универсальным. Тогда как из 11 маркированных включающих местоимение *ta* ‘что’ конкретно-референтным статусом обладает 10, а универсальным – 1. Таким образом, с одной стороны, мы еще раз подтверждаем гипотезу, изложенную выше, что референциальный статус ИГ является одним из факторов лицензоров асимметричного объектного маркирования в современном иврите. А с другой стороны, выдвигаем предположение о наличии дополнительного параметра, который мог бы объяснить, почему, согласно корпусным данным, конструкции, включающие местоимение *mi* ‘кто’, маркируются в роли прямого объекта в несколько раз чаще, чем аналогичные по структуре, статусу определенности и референциальному статусу конструкции, включающие местоимение *ta* ‘что’ [9].

По данным исследовательского корпуса, и особенно, по сопоставлению данных для конструкций *kol mi še-* и *kol ta še-*, мы видим, что носители в некоторых случаях по-разному оформляют прямое дополнение, выраженное ИГ с квантификатором *kol*. Именные группы с квантификатором *kol*, имеющих конкретно-референтный статус, оформляются в подавляющем большинстве случаев универсально (99 %) маркером прямого объекта, тогда как конструкции *kol mi še-* и *kol ta še-* имеют значительную вариативность в маркировании (100 % и 55 % соответственно). Анализ показывает, что существенную роль в лицензировании объектного маркирования такого рода конструкций играет референциальный статус ИГ, кодирующей соответствующий референт, и прагматический контекст высказывания, позволяющий точнее выделить релевантные для ситуации характеристики.

ЛИТЕРАТУРА

1. Glinert L. The Grammar of Modern Hebrew. Cambridge: Cambridge University Press, 1989. 612 p.
2. Givón T. Definiteness and referentiality // Universals of human language / Greenberg J. H., Ferguson C. A., Moravcsik E. A. (eds.). Vol. 4. Stanford: Stanford University Press, 1978. P. 291–330.
3. Gundel J. K., Hedberg N., Zacharski R. Cognitive Status and the Form of Referring Expressions in Discourse // Language. 1993. Vol. 69(2). P. 274–307.
4. Lambrecht K. Information structure and sentence form: Topic, focus, and the mental representations of discourse referents. Cambridge: Cambridge University Press. 1994. 388 p.
5. Heusinger K., Kornfilt J. Partitivity and case marking in Turkish and related languages // Glossa: a journal of general linguistics. 2017. Vol. 2(1). P. 1–40.
6. Сердобольская Н. В., Толдова С. Ю. Дифференцированное маркирование прямого дополнения в финно-угорских языках // Финно-угорские языки: фрагменты грамматического описания. Формальный и функциональный подходы. М.: Языки славянских культур. 2012. с. 59–142.
7. Шмелев А. Д. Русский язык и внеязыковая действительность. М. : ЯСК. 2002. 496 с.
8. Падучева Е. В. Высказывание и его соотнесённость с действительностью. М.: Наука. 1985. 293 с.
9. Алексеева М. Е. Одушевленность и референциальный статус как факторы асимметричного объектного маркирования в современном иврите (на примере вопросительных и относительных местоимений) // Litera. 2023. № 6. С. 210–220.

УДК 811.112'373:811.112'25

Богданова Наталия Альбертовна

кандидат филологических наук,
доцент кафедры фонетики
и грамматики немецкого языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Natallia Bahdanava

PhD in Philology,
Associate Professor of the
Department of Phonetics and
Grammar of the German Language
Belarusian State University
of Foreign Languages
Minsk, Belarus
albertowna@mail.ru

СЕМАНТИЧЕСКАЯ ПРОСОДИЯ НЕМЕЦКИХ ГЛАГОЛОВ СО ЗНАЧЕНИЕМ ПРИЧИННОСТИ

В статье рассматривается явление семантической просодии – особой тональной окраски слова, вызванной коллокационным поведением узлового слова. С помощью метрик, используемых в корпусах текстов, исследуется семантическая просодия немецких глаголов-синонимов с каузальным значением. На основании тональности коллокатов, метрики logDice и частоты коллокации определяется индекс семантической просодии для каждого глагола (SPI). Проводится сопоставительный анализ тональности и семантики коллокатов и SPI для глаголов со значением причинности и выявляются их семантические коллокационные предпочтения. Далее проводится анализ словарных дефиниций глаголов со значением причинности и словарных примеров употребления глаголов с целью выяснить степень репрезентации в словарных дефинициях семантической просодии глаголов, а также проанализировать представленность самых частотных коллокатов в примерах употребления глаголов.

Ключевые слова: семантическая просодия; коллокация; узловое слово; коллокат; тональность.

SEMANTIC PROSODY OF GERMAN VERBS WITH THE MEANING OF CAUSALITY

The article examines the phenomenon of semantic prosody – a special tonal coloring of a word caused by the collocation behavior of a key word. Using metrics employed in text corpora, the semantic prosody of German synonymous verbs with a causal meaning is researched. Based on the tonality of collocations, the logDice metric and collocation frequency, a semantic prosody index (SPI) is calculated for each verb. A comparative analysis of the sentiment and semantics of collocates and the SPI is conducted for causative verbs, revealing their semantic collocational preferences. Furthermore, the study analyzes dictionary definitions of causative verbs and usage examples to determine the extent to which semantic prosody is reflected in their lexical definitions, as well as to assess the representation of the most frequent collocates in verb usage examples.

Key words: semantic prosody; collocation; key word; collocative; tonality.

Семантическая просодия – это лингвистический феномен, отражающий устойчивую тональную окраску слова, формируемую через его типичные сочетания с другими лексическими единицами. В отличие фиксируемого в

словарях денотативного значения лексической единицы, семантическая просодия слова проявляется на уровне дискурса и определяется частотными сильными коллокациями (словами-спутниками) узлового слова.

Термин «семантическая просодия» (*semantic prosody*) (далее – СП) был предложен британским лингвистом, специалистом в области корпусной лингвистики, дискурс-анализа, лексикографии и преподавания английского языка Джоном Синклером в 1990-х годах в рамках исследований по корпусной лингвистике на базе английского языка. Дж. Синклер показал, что значение слова нельзя изучать изолированно, так как оно конструируется через взаимодействие с контекстом. В своих работах Дж. Синклер обосновывает идиоматический принцип организации языка, который заключается в том, что реальная речь состоит из готовых блоков с фиксированной семантикой, чем и обеспечивается беглость устной речи. Например, глагол “set in” – это не просто сумма конститuentов “set + in”, а цельная единица с собственной СП, обозначающая, как правило, негативное начало какого-либо процесса. Слова в словосочетаниях не выбираются изолированно, их значение и грамматика определяются привычными паттернами, которые именно текстовые корпуса делают видимыми. Для Дж. Синклера явление СП – следствие идиоматической природы языка: слова “притягивают” определенные семантические поля, создавая скрытые слои смысла [4].

Идеи Дж. Синклера получили в работах его учеников и последователей свое дальнейшее развитие и были апробированы в основном на базе корпусов английского языка. Так, британский лингвист М. Стаббс указывает на зависимость СП от типа дискурса и предлагает термин “дискурсивная просодия” (*discourse prosody*), подчеркивая, что тональная оценка возникает в социальном контексте, а не только в семантике самого слова [6]. СП слов может влиять на наши суждения и решения, несмотря на отсутствие явной тональности (положительности/ отрицательности). СП действует как тонкий, но мощный фрейм, направляющий рассуждения и решения [3]. СП может быть присуща не лемме в целом, а конкретным грамматическим формам. Например, инфинитив глагола *undergo* и формы настоящего времени гораздо чаще, чем причастия прошедшего времени, встречаются в контекстах, описывающих будущие, гипотетические или нежелательные процедуры/ изменения (*operation, therapy*), часто с оттенком трудности или необходимости. Формы прошедшего времени *underwent, undergone* чаще описывают уже случившиеся события, которые могли быть и нейтральными или даже позитивными (*renovation, transformation, development*). Таким образом, утверждать, что лемма имеет сильную негативную СП, некорректно. Необходимо учитывать грамматическую форму и контекст употребления слова [6].

В работе американского исследователя Дж. Фьюка показана ценность преподавания СП для развития коммуникативной компетенции у изучающих английский как иностранный. Двуязычные словари часто неадекватно отражают или полностью игнорируют СП, что ведет к неправильному употре-

лению слов [2]. Корпусной подход и учет семантической просодии слов повлиял на создание современных словарей (например, *Collins COBUILD*), где акцент сместился с «значений слов» на «слова в контексте» [4].

К ключевым характеристикам СП можно отнести следующие:

1) **Имплицитность.** СП слов чаще всего не закреплена в словарях и может не осознаваться носителями языка. Например, русское слово «последствие» не имеет явной негативной тональности, но если обратиться к корпусу текстов и посмотреть на коллокации слова (например, коллокаты-прилагательные), то можно увидеть, что существительное чаще и активнее сочетается с прилагательными отрицательной тональности. При этом, если проанализировать прилагательные-коллокаты к лемме “последствие” в единственном числе, отсортированные по значению метрики *logDice* (9–8), то среди них будут как тонально отрицательные, так и нейтральные прилагательные: *неизбежный, неминуемый, естественный, логический, непременный, роковой, прямой, негативный*. У леммы “последствие” во множественном числе картина совсем иная: *logDice* начинается от 10 (то есть усиливается коллокационная связь с узловым словом) и коллокаты со значением *logDice* 10-9 выглядят почти сплошь негативно: *негативный, чреватый, катастрофический, вредный, пагубный, печальный*. В работе последователя Дж. Синклера британского лингвиста Д. Стюарта показано, что СП может быть скрыта от интуиции носителей, что, однако, не означает, что она скрыта всегда. Д. Стюарт предполагает, что для некоторых слов с очень сильной и последовательной СП (*commit*) носители языка могут осознавать негативную тональность. «Скрытость» не должна быть обязательным и универсальным критерием СП слова. Это скорее потенциальная характеристика, проявляющаяся по-разному для разных слов [5].

2) **Контекстуальная зависимость.** СП формируется через повторяющиеся паттерны употребления слова. СП контекстуально обусловлена: если СП слова негативная, то в сочетаниях со словами нейтральной или положительной тональности контекст может раскрывать негативные аспекты. В статье М. В. Сергеевой и В. В. Денисовой доказано, что прилагательное *global* в современном англоязычном дискурсе (на материале СОСА за 2010 г.) обладает устойчивой негативной семантической просодией. Оно преимущественно сочетается с лексемами, обозначающими проблемы, угрозы и кризисы. Даже в сочетаниях с нейтральными (*global economy, global market*) или позитивными лексемами (*global stability, global justice*) контекст чаще всего раскрывает негативные аспекты (например, *global economy* описывается в ситуациях кризиса или нестабильности) [1].

3) **Коллокационная предсказуемость.** Слова-коллокаты узлового слова тяготеют к определенным семантическим группам. М. Стаббс говорит в этой связи о «семантической предрасположенности» (*semantic preference*), то есть связи слова с определенной семантической группой. Например, узловое слово *подвергаться* притягивает определенные семантические группы коллокатов (негативные воздействия: *критика, нападки, пытки, преследования*

и т.д.) Такие сочетания стабильны и встречаются в корпусах чаще, чем случайные комбинации. Это явление помогает предсказывать, с какими семантическими группами слов будет сочетаться узловое слово, что крайне важно для обеспечения адекватного перевода текстов, для обучения языкам и для создания словарей (например, в *Collins Cobuild* такие сочетания выделяются специально).

По тональной направленности традиционно выделяют:

- негативную/отрицательную СП: узловые слова, тяготеющие к сочетаниям с коллокатами отрицательной тональности;
- позитивную/положительную СП: при преобладании слов-коллокатов, ассоциирующихся с успехом или одобрением, желательными событиями или состояниями;
- нейтральную СП: отсутствие явной преобладающей тональности коллокатов.

Хотя исследовательская традиция СП еще не сложилась, СП чаще определяют для тонально нейтральных узловых слов, в то время как слова с явной выраженной оценкой редко анализируются на предмет СП, хотя их коллокационные профили тоже системны. В этой связи часто возникает путаница между понятиями «семантическая просодия» слова и «коннотация» слова. Это различные, хотя и взаимосвязанные явления. Коннотация – это часть лексического значения слова вне зависимости от его коллокационного поведения. Коннотацией называют дополнительное, устойчивое эмоциональное, оценочное или стилистическое значение, прикрепленное к слову в сознании носителей языка и часто фиксируемое в словарях. СП в свою очередь – следствие именно коллокационного поведения слова, напрямую зависящее от его окружения. Слово с нейтральной тональностью может иметь ярко выраженную (часто негативную) СП (*cause, commit*). И наоборот, слово с коннотативным значением (*пилить*) может не иметь специфической СП, ограничивающей его сочетаемость только с позитивным или негативным тональным окружением. Таким образом, СП – это свойство употребления слова, устойчивая тенденция слова сочетаться с окружением определенной эмоциональной окраски (позитивной, негативной, нейтральной).

Для нашего исследования мы обратились к группе глаголов со значением причинности, отобранных из словаря DWDS и представленных в нем как синонимы: *verursachen, hervorrufen, auslösen, bewirken, erwecken, führen zu D., sorgen für Akk.* (в соответствующих значениях). На этапе сбора корпусных данных был использован корпус текстов DWDS. Для анализа СП слов использовались данные из раздела «Профиль слова» (Wortprofil) корпуса DWDS. Текущая версия раздела (2025 г.) опирается на корпуса общим объемом ~13 млн текстов, включающих 283 млн предложений и 7 млрд слов. Профили слов доступны для 400 000 лемм и содержат в сумме 22,5 млн уникальных коллокаций. Эмпирическую базу составляют признанные общенациональные актуальные немецкоязычные газеты (ежедневные и ежене-

дельные), а также тексты из «Базового корпуса» (художественная и научно-популярная литература, научные работы, пресса), охватывающие период с начала XX века. Важно, что в профиле слова отображаются только статистически значимые коллокации, удовлетворяющие критериям минимальной частоты ($F \geq 5$) и значению меры ассоциации $\logDice > 0$.

Для каждого глагола было отобрано по 20 коллокатов, отсортированных по значению $\logDice$. $\logDice$ помогает отфильтровать частые, но слабые связи, которые «зашумляют» картину СП. Коэффициент $\logDice$ отражает статистическую значимость связи между словами. Чем выше значение, тем устойчивее коллокация. Выделяют сильную связь ($\logDice \geq 7.0$) – идиоматичные, высокочастотные сочетания; умеренную связь ($5.0 \leq \logDice < 7.0$) – типичные, но вариативные сочетания; слабую связь ($\logDice < 5.0$) – редкие или контекстно-зависимые коллокации. В таблице 1 представлено распределение первых 20 коллокатов по силе связи с исследуемыми узловыми словами:

Т а б л и ц а 1

Распределение коллокатов
по силе связи с узловыми словами

Глагол	Сильная связь (≥ 7.0)/F	Умеренная связь ($5.0-6.9$)/F	Слабая связь (< 5.0)/F
verursachen	Schaden (9.4/12414), Unfall (9.0/7141), Kosten (8.7/12760), Verkehrsunfall (7.7/1381), Sachschaden (7.5/1177), Mehrkosten (7.5/1185), Lärm (7.3/1269), Schmerz (7.2/1504), Aufregung (7.1/1135), Wirbel (7.1/893)	Brand (6.9/1218), Krebs (6.7/836), Stau (6.7/775), Leid (6.6/777), Millionenschaden (6.3/440), Chaos (6.2/688), Emission (6.2/549), Elfmeter (6.2/497), Ärger (6.1/694), Störung (6.1/548)	—
hervorrufen	Empörung (8.5/1676), Reaktion (8.3/3654), Kopfschütteln (7.9/669), Unmut (7.6/874), Assoziation (7.6/641), Erstaunen (7.3/489), Echo (7.3/628), Entsetzen (7.3/552), Verwunderung (7.1/419), Besorgnis (7.1/497), Protest (7.1/1923), Irritation (7.0/443)	Begeisterung (6.9/708), Bestürzung (6.8/293), Unverständnis (6.7/356), Widerspruch (6.7/854), Stirnrunzeln (6.6/257), Kritik (6.6/2561), Entzündung (6.6/285), Emotion (6.5/454)	—

bewirken	Gegenteil (8.4/3968), Veränderung (7.2/1856), Wunder (7.2/1531), Umdenken (7.2/595), Positiv (7.1/573), Gut (7.0/1454)	Sinneswandel (6.0/226), Wandel (6.0/529), Verhaltensänderung (5.9/205), Groß (5.8/254), Änderung (5.7/559), Verbesserung (5.6/443), Umschwung (5.5/155), Rückgang (5.4/369), Verschiebung (5.2/186), Effekt (5.1/275), Stimmungsumschwung (5.1/117), Anstieg (5.1/321), Meinungsumschwung (5.1/109), Wende (5.0/266)	—
auslösen	Empörung (8.7/5572), Debatte (8.5/9272), Reaktion (8.4/6938), Alarm (8.3/3916), Diskussion (7.9/6512), Welle (7.9/3812), Protest (7.8/5181), Kontroverse (7.8/2739), Begeisterung (7.8/3180), Unruhe (7.6/2904), Entsetzen (7.6/2338), Kettenreaktion (7.4/1970), Irritation (7.4/1991), Sturm (7.4/2627), Panik (7.3/2139), Kopfschütteln (7.3/1800), Echo (7.1/1749), Besorgnis (7.1/1673)	Brand (6.9/1810), Skandal (6.8/1833)	—
erwecken	Eindruck (11.1/36719), Anschein (10.6/5959),	Fossil (6.9/347) Mitleid (6.3/285), Misstrauen (6.2/372), Assoziation (6.2/238), Verdacht (6.0/719), Neid (6.0/210), Argwohn (5.9/168), Aufmerksamkeit (5.8/483), Illusion (5.6/246), Vertrauen (5.6/572), Sympathie (5.3/194), Neugier (5.2/133)	Gefühl (4.8/541), Tote (4.8/308), Interesse (4.7/823), Schein (4.7/121), Hoffnung (4.7/519), Bewunderung (4.7/85)

führen zu	Erfolg (7.2/16154), Ergebnis (7.0/14128),	Tod (6.3/7955), Verlust (6.3/7108), Ende (6.0/11168), Anstieg (5.9/4646), Sieg (5.7/4973), Konflikt (5.7/4524), Rückgang (5.7/4119), Problem (5.7/6443), Ziel (5.6/5677), Belastung (5.5/3674), Spannung (5.5/3586), Veränderung (5.4/3627), Protest (5.4/3529), Diskussion (5.3/3563), Verzögerung (5.2/2851), Verbesserung (5.2/2903), Auseinandersetzung (5.0/2717), Situation (5.0/3069)	—
sorgen für	Aufsehen (9.8/26567), Aufregung (9.0/15608), Schlagzeile (8.9/15414), Unruhe (8.5/11850), Furore (8.3/9305), Stimmung (8.2/11452), Wirbel (8.2/8455), Unmut (7.8/6845), Überraschung (7.8/7427), Sicherheit (7.8/10439), Ärger (7.7/7022), Verwirrung (7.6/5635), Empörung (7.5/5514), Irritation (7.3/4797), Unterhaltung (7.2/4679), Ordnung (7.2/5545), Gesprächsstoff (7.1/3956), Diskussion (7.1/6106), Ruhe (7.0/4742)	Spannung (6.9/4050)	—

Как видно из таблицы, глаголы *hervorrufen*, *auslösen*, *sorgen für* обладают максимальным числом коллокатов с сильными связями, глаголы *bewirken* и *führen zu* имеют в основном умеренные связи с коллокатами, у глагола *verursachen* коллокаты поровну распределены между сильными и умеренными связями, у глагола *erwecken* связи скорее распределены между умеренными и слабыми, однако именно этот глагол имеет самый сильно связанный коллокат *Eindruck* (11.1). Уже по данным таблицы можно предвари-

тельно судить о тональном «коллокационном предпочтении» или о его отсутствии у анализируемых глаголов. Например, глаголы *verursachen* и *hervorrufen* сочетаются с коллокатами негативной тональности (*Schaden, Unfall, Kosten, Verkehrsunfall, Sachschaden, Mehrkosten, Lärm, Schmerz, Aufregung, Wirbel* и *Empörung, Kopfschütteln, Unmut, Entsetzen, Protest, Irritation* соответственно), глагол *bewirken* сочетается с коллокатами скорее позитивной или нейтральной тональности (*Gegenteil, Wunder, Positiv, Gut, Veränderung, Umdenken*). Однако высокое значение logDice само по себе еще не говорит о преобладающей СП узлового слова. Именно в корреляции с высокой частотностью коллокации такая связь показательна для СП. Например, показательными являются связи с леммами *Schaden, Kosten* для глагола *verursachen*, лемма *Eindruck* для глагола *erwecken*, леммы *Aufsehen, Aufregung* для глагола *sorgen für*. В таблице есть несколько примеров, где коллокаты имеют умеренный показатель LogDice, но при этом низкую частотность. Это означает, что сочетание глагола с данным существительным является устойчивым и семантически значимым, но встречается в текстах относительно редко (*erwecken + Fossil, bewirken + Sinneswandel*). Для объективации картины СП мы определяем учесть обе метрики и рассчитать индекс СП глаголов со значением причинности (SPI) по формуле

$$SPI(i) = \sum (S(i) * L(i) * \log_{10}(F(i)) / \sum (L(i) * \log_{10}(F(i))), \text{ где}$$

1. S(i) Тональность коллоката по шкале “хорошо” – “нейтрально” – “плохо” со значениями -1, 0, +1 соответственно;
2. L(i) – logDice коллоката;
3. F(i) – частотность заданной коллокации.

Результаты расчетов по исследуемым узловым словам приведены в таблице 2.

Т а б л и ц а 2

Семантическая просодия немецких глаголов
со значением причинности

Ранг	Глагол	SPI	Семантическая Просодия	Основные семантические преференции коллокатов
1	verursachen	–1	Ярко выраженная Отрицательная	Негативные последствия и события: ущерб, аварии, издержки, боль, болезнь, хаос, помехи.
2	hervorrufen	–0,59	Отрицательная	Преимущественно негативные реак- ции: возмущение, недовольство, критика, ужас, протест
3	auslösen	–0,53	Отрицательная	Частотные негативные эмоциональ- ные реакции: возмущение, протест, тревога, паника, скандал, споры, беспокойство

4	führen zu D.	–0,17	Нейтрально-Отрицательная	Широкий спектр последствий: успех, результат, но часто негативный исход – смерть, потеря, конфликт, проблема.
5	sorgen für Akk.	–0,15	Нейтрально-отрицательная	Состояния/Условия: сенсация, безопасность, порядок, волнение, раздражение.
6	erwecken	0,06	Нейтральная	Реакции разного рода: впечатление, чувство, видимость, интерес, недоверие, подозрение, жалость.
7	bewirken	0,25	Положительная	Позитивные преобразования: позитивные изменения, улучшения, чудеса, добро.

Сопоставление данных таблицы с данными словарных дефиниций узловых слов в DWDS показало, что в словарных статьях содержится лакунарное отражение и даже полное игнорирование СП слов, хотя в примерах могут быть отражены слова тональности преобладающих коллокатов с сильной связью, но это не всегда прослеживается системно. Например, дефиниция глагола *verursachen* «*die Ursache von etwas sein, etwas bewirken*» нейтральна. Но мы определили глагол *verursachen* как глагол с яркой отрицательной СП, в дефиниции же употреблен глагол с положительной СП (*bewirken*). В примерах, однако, представлены сочетания с существительными исключительно негативной тональности: *Schaden, Unfall, Kosten, Lärm*, что в целом соответствует самым сильным связям узлового слова. Дефиниция глагола *hervorrufen* «*etw. verursachen, bewirken, auslösen*» также нейтральна, в примерах сочетаний, однако, приведены не самые репрезентативные по тональности коллокаты: *Begeisterung, Eindruck, Entzündung* (а по данным профиля слова это *Empörung* (8.9), *Reaktion* (8.3), *Kopfschütteln* (7.9), *Unmut* (7.6), *Entsetzen* (7.3)), в то время как коллокаты позитивной тональности (*Begeisterung* (6.9)) занимают более низкие позиции по силе связи с узловым словом. Дефиниция игнорирует доминирующую отрицательную СП слова и примеры скорее не репрезентативны. Из всех проанализированных глаголов лучший пример представляет собой описание глагола *führen zu*, чья дефиниция также нейтральна, а примеры сочетаний: *Ergebnis, Chaos, Verzweiflung* (Коллокаты: *Erfolg* (7.2), *Ergebnis* (7), *Tod* (6.3), *Verlust* (6.3)) отражают двойственность тональности коллокатов.

Методика определения СП слова пока содержит много спорных моментов: учет количества коллокатов, учет различных корпусных метрик, определения границ языковых элементов для изучения СП, во многом субъективное определение тональности коллокатов, ширина дискурсивного контекста и др. Однако большинство исследователей СП едины в понимании этого феномена как ключевого элемента идиоматичности языка, связывающего лексику с культурными и прагматическими установками носителей языка. Изучение

СП необходимо для избежания калькирования негативной тональности слова при переводе, для обучения естественному употреблению слов через самые репрезентативные коллокации и с учетом тональности переводных эквивалентов родного и изучаемого языка, а также для разработки алгоритмов сентимент-анализа фраз и текстов в компьютерной лингвистике. Понимание семантической просодии расширяет представление о языке как о системе, детализирует и эксплицирует механизмы формирования значения слова через сеть взаимосвязанных контекстов.

ЛИТЕРАТУРА

1. Сергеева М. В., Денисова В. В. Изучение семантической просодии слова на базе анализа корпусов // Вестник Нижегородского государственного лингвистического университета им. Н. А. Добролюбова. Серия: Языкознание. 2012. № 10 (697). С. 86–91. URL: <https://elibrary.ru/item.asp?id=17778203> (дата обращения: 05.06.2025).
2. Fuqua J. Semantic Prosody: The Phenomenon of "Prosody" in Lexical Patterning // The Journal of Language Teaching and Learning. 2014. № 2. P. 76–83. URL: https://www.academia.edu/7651384/Semantic_Prosody_The_Phenomenon_of_Prosody_in_Lexical_Patterning (дата обращения: 05.06.2025).
3. Hauser D., Schwarz N. Semantic Prosody: How Neutral Words With Collocational Positivity/Negativity Color Evaluative Judgments // Journal of Experimental Psychology: General. 2023. URL: https://www.researchgate.net/publication/368725854_Semantic_Prosody_How_Neutral_Words_With_Collocational_Positivity_Negativity_Color_Evaluative_Judgments (дата обращения: 05.06.2025).
4. Sinclair J. Corpus, Concordance, Collocation. Oxford : Oxford University Press, 1991. 179 p.
5. Stewart D. Semantic Prosody: A Critical Evaluation. London : Taylor & Francis, 2010. 193 p.
6. Stubbs M. Words and Phrases: Corpus Studies of Lexical Semantics. Oxford : Blackwell Publishing, 2001. 287 p.

УДК 81-11

Будникова Елена Игоревна

кандидат филологических наук, доцент,
заведующий кафедрой лексикологии
и стилистики английского языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Alyona Budnikova

PhD in Philology, Associate Professor,
Head of the Department of English
Lexicology and Stylistics
Belarusian State University
of Foreign Languages
Minsk, Belarus
geronimo1981@mail.ru

СЕМАСИОЛОГИЧЕСКИЕ И ОНОМАСИОЛОГИЧЕСКИЕ ПРИНЦИПЫ ОТБОРА ЛЕКСИЧЕСКОГО МАТЕРИАЛА ДЛЯ УЧЕБНЫХ ЦЕЛЕЙ: СЛОВАРЬ VS. КОРПУС

В статье описывается методика отбора лексических единиц в качестве базы для языковых и условно-речевых упражнений на основании сочетания двух подходов к анализу лексической семантики – от формы к ее значению (семасиологии) и от значения к форме его выражения (ономасиологии) – и, соответственно, сочетания двух источников материала – словарей (а также лексических баз) и корпуса текстов. Применение методики демонстрируется на материале прилагательных, описывающих вкусовые характеристики блюд в английском языке. Анализируется эффективность предложенного подхода.

Ключевые слова: корпус текстов; лексико-семантическое поле; тематическая лексика; семасиология; ономасиология.

SEMASIOLOGICAL AND ONOMASIOLOGICAL PRINCIPLES IN CHOOSING LEXICAL ITEMS FOR STUDY PURPOSES: DICTIONARY VS. CORPUS

The article describes the methodology of choosing vocabulary units as bases for further language and speech exercises based on a combination of two principles of semantic analysis – from form to meaning (semasiology) and from meaning to form (onomasiology) – and, consequently, on a combination of two major sources of material – dictionaries (including lexical bases) and corpora. The methodology is tested on a case study of English adjectives describing taste of dishes. Its efficiency is analyzed.

Key words: corpora; lexical field; semantic field; thematic vocabulary; semasiology; onomasiology.

Отбор лексических единиц для включения в тематический словарь или в качестве ресурсной базы для составления языковых и речевых упражнений подчиняется ряду принципов, широко обсуждаемых в научной литературе (см. работы Л. В. Щербы, И. В. Рахманова, Н. В. Николаева, Н. Д. Гальсковой, Н. И. Гез, И. М. Берман, С. И. Белотеловой, Д. М. Усмановой и др.). Как правило, говорят об общедидактических принципах, которые предполагают научность и системность в отборе единиц, последовательность в предъявлении и доступность материала, учет возрастных особенностей и активности субъекта, а также новизну отобранных слов (см. обзор в [1]). В качестве методических рекомендаций по отбору тематической лексики могут приво-

дятся принципы тематической ценности (насколько единицы востребованы при выполнении коммуникативных упражнений на заданную тему), резервности (потенциальная значимость единиц для последующих тем и ситуаций общения), легкости для запоминания и пр. Наконец, среди собственно лингвистических принципов выделяются такие параметры единиц, как частотность (не всегда коррелирует с коммуникативной значимостью), вхождение в гиперо-гипонимическую иерархию (единицы т.н. базового уровня категоризации являются, как правило, более востребованными), многозначность (в т.н. основной словарный фонд входят преимущественно многозначные слова), учет словообразовательных и сочетаемостных возможностей слова (единицы с развитыми словообразовательными и сочетаемостными связями являются более ценными для запоминания) и др. При этом, закономерно, что в зависимости от целей урока, категории обучающихся и типов упражнений значимость лингвистических принципов может варьироваться (см., например, [2; 3]).

Целью настоящей статьи является обоснование методики отбора тематической лексики для учебных целей на основании комбинации семасиологического и ономасиологического подходов и с использованием возможностей корпусной лексикографии. При этом думается, что отобранный материал далее может быть скорректирован под особенности той или иной группы обучающихся, вид учебного занятия, тип формируемой компетенции и т. д.

Общеизвестно, что семасиология как одна из сторон лексической семантики изучает, какие значения может принимать та или иная языковая форма. В рамках семасиологического направления выявляется типология лексических значений, связи между разными значениями одной формы, классифицируются компоненты значения слова и т. д. Соответственно, семасиологический принцип отбора единиц для включения в тематический словарь предполагает, что слова отбираются в зависимости от их значений, а именно – ключевых слов или классификаторов в их словарных дефинициях. Это системный подход, ограничивающийся значением единицы вне какого-либо контекста употребления.

В свою очередь ономасиология исходит из того, какими языковыми формами может быть выражена та или иная мысль или денотат, какие аспекты обозначаемой сущности отражены в этих формах, какие отношения связывают их и т. д. Ономасиологический принцип отбора единиц предполагает, что слова отбираются в зависимости от того, используются ли они для указания на то или иное значение. Это внесистемный, контекстуальный подход, в результате которого в перечень тематической лексики попадают не только т.н. ядерные, но и периферийные единицы.

Для иллюстрации того, как работает предложенная методика и как полученный реестр тематической лексики может быть в дальнейшем использован, рассмотрим этапы построения лексико-семантического поля прилагательных вкуса в английском языке, которые коррелируют с указанными ранее критериями отбора.

Предположим, что для учебных целей требуется выявить реестр лексических единиц, которые можно использовать в английском языке для указания на дегустационную характеристику пищи (слова, описывающие вкусовые качества). Первым этапом отбора является установление круга имен прилагательных¹ в английском языке, в значении которых зафиксирован семантический компонент вкуса, выраженный словами-идентификаторами *taste* ‘вкус’, *flavor* ‘аромат’ и *smell* ‘запах’, что логически следует из физиологической близости указанных ощущений и зафиксировано в словарных дефинициях данных единиц. Отбор прилагательных осуществлялся по онлайн-версиям толковых словарей с опцией тезаурусного поиска [4; 5; 6].

В реестр исследуемых единиц, в соответствии с принципами отбора, попадают, прежде всего, производные² слов-идентификаторов: *tasty* ‘вкусный’, *tasteless* ‘безвкусный’, *flavoursome* ‘вкусный’, *flavourless* ‘безвкусный’, *smelly* ‘дурно пахнущий’. В целом же в ходе первого этапа отобраны 29 единиц, среди которых можно выделить следующие группы:

1) ядерные прилагательные, описывающие наличие или отсутствие вкуса (часто с оценочными компонентами значений и производные от ключевых слов): *tasty* ‘вкусный’, *tasteless* ‘безвкусный’, *flavoursome* ‘вкусный’, *flavourless* ‘безвкусный’, *smelly* ‘дурно пахнущий’, *delicious* ‘очень вкусный’;

2) прилагательные, указывающие на конкретный вкус продукта (иногда производные от наименования этого продукта): *sweet* ‘сладкий’, *sour* ‘кислый’, *salty* ‘соленый’, *bitter* ‘горький’, *vinegary* ‘уксусный’ и пр.;

3) прилагательные, синонимичные указанным выше единицам и определяющиеся с их помощью в словарях: *brackish* ‘somewhat salty’, *mouthwatering* ‘delicious or appealing’, *stale* ‘tasteless or unpalatable from age’ и пр.

На втором этапе выявляются лексические единицы, употребляемые для описания вкусовых качеств в речи. Источником материала выступает корпус текстов³, а маркером прилагательного в исследуемом значении – атрибутивная позиция при ключевых словах *taste*, *flavor*, *smell*. Общее количество отобранных контекстов составило 50 сочетаний с прилагательными⁴.

Среди отобранных на основании ономаσιологических критериев единиц можно выделить следующие группы:

¹ Качество, или непроцессуальный признак предмета, является категориальной характеристикой прилагательного как части речи, поэтому в исследуемый реестр попали, прежде всего имена прилагательные.

² Далее в исследовании мы рассматривали морфологические производные прилагательных вкуса только в случае их фиксации словарями или в контекстном использовании в качестве модификаторов единиц *taste*, *smell*, *flavour*. Результаты деривации исследуемых прилагательных по регулярным моделям словообразования, не отмеченные в словарях как отдельные единицы, не рассматривались.

³ Для отбора контекстов использовался Британский национальный корпус (BNC) [7].

⁴ При этом в реестр вошли только контексты, в которых описываются те или иные характеристики пищи или блюд. Описания с помощью указанных ключевых слов иных ситуаций не рассматривались в качестве релевантных контекстов.

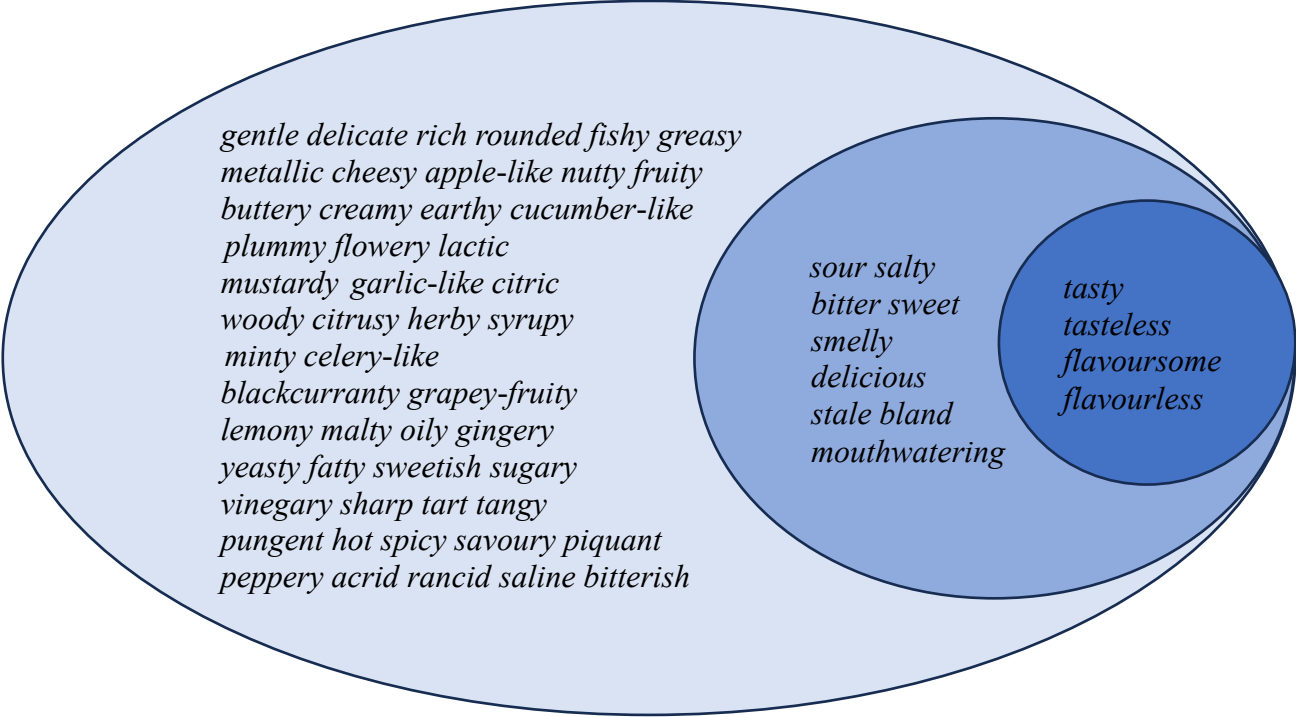
1) прилагательные широкой оценочной семантики (*exotic* ‘экзотический’, *satisfying* ‘удовлетворяющий’), ср. *blandly-flavoured foods reduced the strong and exotic taste of many herbs* [7] ‘продукты с приторным вкусом уменьшают сильный и экзотический вкус многих трав’;

2) прилагательные с частнооценочной семантикой, которые описывают не столько вкусовые характеристики пищи, сколько физиологическую реакцию человека на пищу (*appetizing* ‘вызывающий аппетит’, *nauseating* ‘вызывающий тошноту или отвращение’), ср. *the dog began to produce saliva in response to the appetizing smell* [7] ‘собака начала выделять слюну в ответ на аппетитный запах’;

3) прилагательные с частнооценочной семантикой, которые указывают на ряд дополнительных свойств пищи, таких как способность к насыщению (*light* ‘легкий’, *hearty* ‘сытный, плотный’, *full* ‘полный, насыщенный’), температуру (*warm* ‘теплый’, *cool* ‘холодный’, *fresh* ‘свежий’, *refreshing* ‘освежающий’), отсутствие примесей (*clean* ‘чистый’, *pure* ‘чистый’, *natural* ‘натуральный’, *dry* ‘сухой’), резкость и интенсивность (*smooth* ‘мягкий’, *mild* ‘мягкий’, *soft* ‘мягкий’, *gentle* ‘нежный’, *delicate* ‘тонкий, изысканный’, *rich* ‘богатый, насыщенный’);

4) прилагательные, производные от существительных-наименований прототипа вкусовых качеств (*fishy* ‘рыбный’, *greasy* ‘жирный’, *metallic* ‘металлический’, *cheesy* ‘сырный’, *apple-like* ‘яблочный’ и т.д.).

Таким образом, на основании примененной методики построено лексико-семантическое поле прилагательных вкуса в английском языке, включающее ядро, ближнюю и дальнюю периферию (рисунок).



Ядро, ближняя и дальняя периферия
лексико-семантического поля наименований вкусовых качеств в английском языке

В качестве преимуществ отобранного по предложенной методике тематического вокабуляра можно отметить следующие:

1) наличие производных единиц, понимание которых не представляет сложностей при среднем и средне-продвинутом уровне владения английским языком (*vinegary, spicy, citrusy, herby, fishy, etc.*),

2) наличие единиц, иллюстрирующих словообразовательные модели, что способствует их лучшему усвоению и развитию аналитических навыков (*fatty, garlic-like, rounded, sweetish*),

3) наличие синонимичных и антонимичных единиц, что позволяет выстроить соответствующие ряды и существенно расширить словарный запас обучающихся (ряд синонимов «кислый»: *sour, vinegary, rancid, acrid*; ряд синонимов «сладкий»: *sweet, sugary, sweetish, syrupy*; ряд синонимов «острый»: *hot, sharp, spicy, piquant, pungent, peppery*; ряд синонимов «вкусный»: *appetizing, mouthwatering, satisfying, delicious*; ряд синонимов «нежный»: *smooth, mild, soft, gentle, delicate*; антонимичные пары: *bland* ‘пресный’ – *tangy* ‘острый’, *light* ‘легкий’ – *hearty* ‘сытный, плотный’, etc.).

Отобранный языковой материал может быть использован в качестве базы для языковых и речевых упражнений. Языковые упражнения включают задания на подбор синонимов и антонимов из числа предложенных вариантов, на словообразование (подбор словообразовательного форманта к предложенной основе). Примерами условно-речевых упражнений с отобранным материалом могут служить следующие типы заданий:

1) Which of the following tastes are pleasant and which are unpleasant to you? Why? When did you last taste it? / Какой из вкусов тебе приятен и какой неприятен? Почему? Когда ты последний раз их пробовал? (Варианты: *greasy, creamy, peppery, sugary, garlic-like, minty, plummy, mustardy, metallic, nutty*).

2) What dishes would you describe as: *nauseating? satisfying? exotic? bland?* / Какие блюда ты бы описал как: *тошнотворные? приятные? экзотические? пресные?*

Применяя указанные ранее методические рекомендации по отбору тематической лексики к отбранному нами вокабуляру, можно сделать следующие выводы:

1) единицы имеют высокую тематическую ценность благодаря разнообразию оттенков описываемого значения;

2) единицы характеризуются потенциальной резервностью, поскольку многие из них многозначны, следовательно, легко переносятся на другие области (цветовые характеристики, оценочное значение, качества характера и пр.);

3) единицы не предполагаются для списочного запоминания, но их прозрачная внутренняя форма во многом способствует их закреплению, по крайней мере, в кратковременной памяти;

4) единицы сами по себе не являются особо ценными с точки зрения коммуникации, однако, легко встраиваются в разговорные формулы на тему «Посещение ресторана», «Приготовление блюда».

Таким образом, соединение семасиологического и ономасиологического подходов к отбору тематического материала означает отбор единиц как по лексикографическим источникам (тезаурусам, толковым словарям, лексическим базам данных), так и по корпусу текстов. В результате отбирается достаточно широкий реестр единиц, из которых можно подбирать лексическую базу для языковых и речевых упражнений. Применение указанной методики отбора целесообразно при работе с обучающимися среднего и продвинутого уровней владения английским языком.

ЛИТЕРАТУРА

1. Касимова Р. Т. Отбор тематической лексики в учебных целях. Вестник РУДН. Серия «Вопросы образования: языки и специальность». 2017. Т. 14. № 4. С. 676–686.
2. Дмитрусенко И. Н. Принципы отбора лексических единиц для формирования рецептивной лексики в процессе самостоятельного чтения при модульной системе обучения. Вестник ЮрГУ. Серия «Образование. Педагогические науки». 2015. Т. 7. № 1. С. 69–72.
3. Пирогова Н. Г. К вопросу отбора лексического материала для чтения и аудирования при обучении студентов иностранному языку. Проблемы современного образования. 2019. № 1. С. 109–117.

ИСТОЧНИКИ ЯЗЫКОВОГО МАТЕРИАЛА

4. The Free Dictionary [Electronic resource]. – Mode of access : <https://www.thefreedictionary.com/>. – Date of access : 26.09.2024.
5. Merriam-Webster Dictionary [Electronic resource]. – Mode of access : <https://www.merriam-webster.com/>. – Date of access : 26.09.2024.
6. Collins Dictionary [Electronic resource]. – Mode of access : <https://www.collinsdictionary.com/>. – Date of access : 26.09.2024.
7. British National Corpus [Electronic resource]. – Mode of access : <http://www.natcorp.ox.ac.uk/>. – Date of access : 28.09.2024.

УДК 81'

Ефимова Анна Дмитриевна
кандидат филологических наук,
доцент кафедры английского языка
Государственный гуманитарно-тех-
нологический университет
г. Орехово-Зуево, Россия

Anna Efimova
Candidate of Philological Sciences,
Associate Professor of the English
Language Department
State University of Humanities
and Technology
Orekhovo-Zuyevo, Russia
Lady-ann2792@yandex.ru

МЕТАФОРИЗАЦИЯ И ЭВФЕМИЗАЦИЯ КОНЦЕПТУАЛЬНОГО ЗАИМСТВОВАНИЯ «ИНВЕСТИЦИЯ» НА ОСНОВЕ АНАЛИЗА НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА

Данная статья направлена на изучение механизмов когнитивно-языкового регулирования концептуального заимствования «инвестиция» в русской лингвокультуре. На основе анализа текстов Национального корпуса русского языка было выявлено, что в процессе эволюции заимствование «инвестиция» прошло процесс детерминологизации. Оно стало употребляться в разных сферах общественной жизни не только для обозначения экономической выгоды, но и по отношению к получению преимуществ в различных других аспектах, в частности, в усилении политического влияния или увеличении превосходства над соперником. Так, в текстах политической сферы основными механизмами когнитивно-языкового освоения заимствования стали метафоризация и эвфемизация, представляющие мощные средства манипулятивного воздействия на общественное сознание, целенаправленно создающие образ врага или образ друга через стратегии сотрудничества или соперничества. В основу метафорической репрезентации лег перенос в сферу межличностных отношений и в военную сферу, использование эмоционально окрашенных слов и выражений. Сделаем вывод о расширении концептуального содержания заимствования «инвестиция» и его актуальности для современной русской лингвокультуры.

Ключевые слова: концептуальное заимствование; инвестиция; глобализация; метафора; эвфемизм, когнитивно-языковое регулирование.

METAPHORIZATION AND EUPHEMIZATION OF THE CONCEPTUAL BORROWING "INVESTMENT" BASED ON THE ANALYSIS OF THE NATIONAL CORPUS OF THE RUSSIAN LANGUAGE

This article aims to study the mechanisms of cognitive-linguistic regulation of the conceptual borrowing "investment" in Russian linguoculture. Based on the analysis of the texts of the National Russian Corpus, it was revealed that in the process of evolution, the borrowing "investment" has undergone a process of determinization. It began to be used in various spheres of public life not only to denote economic benefits, but also to gain advantages in various other aspects, in particular, in strengthening political influence or increasing superiority over a rival. Thus, in the texts of the political sphere, the main mechanisms of cognitive and linguistic

assimilation of borrowing have become metaphorization and euphemization, which represent powerful means of manipulative influence on public consciousness, purposefully creating an image of an enemy or an image of a friend through strategies of cooperation or competition. The metaphorical representation is based on the transfer to the sphere of interpersonal relations and to the military sphere, the use of emotionally colored words and expressions. To conclude the conceptual content of the loan "investment" has expanded and it is relevant for modern Russian linguoculture.

Key words: conceptual borrowing; investment; globalization; metaphor; euphemism, cognitive-linguistic regulation.

В настоящее время в эпоху глобализации происходит заимствование целых пластов лексики в разные сферы общественной жизни. Особенно важными становятся концептуальные заимствования, которые транслируют ценности стран-проводников глобализации в принимающие лингвокультуры и отражают концептосферу отдающей культуры. При этом происходит когнитивно-языковая обработка заимствования, в результате которой под влиянием принимающей лингвокультуры заимствование приобретает специфические образно-ценностные характеристики: «...происходит встраивание содержания <....> заимствования в ее [лингвокультуры] концептуальное пространство, сопоставление с уже существующими элементами концептосферы, приобретение определенных оценочных характеристик» [1, с. 83].

Освоение заимствования и соответствующего концепта, определяется различными факторами когнитивно-языкового регулирования, позволяющими встроить данное заимствование в систему ценностных координат принимающей лингвокультуры. Когнитивно-языковое освоение заимствования может быть обусловлено принадлежностью к определенному типу заимствования: определяемому его целями (В. И. Карасик выделяет игровые, паразитарные, технические, ценностные [2, с. 106]); на основе выраженности этнокультурной специфики и наличия синонимии в принимающей культуре (макаронизмы, разные виды экзотизмов и т. д.); на основе выполняемой функциональной нагрузки (терминологические, компьютерный сленг, общеупотребительная лексика); в зависимости от принадлежности к определенному дискурсу (педагогическому, рекламному, политическому, экономическому и т. д.), что детерминирует этапы освоения заимствования, его трансформации в процессе эволюции, приобретаемые семантические признаки и их социальную оценку. Так, особая природа терминологических концептуальных заимствований обуславливает особенности их адаптации. При этом заимствование обычно имеет более конкретное, узкое значение по сравнению со словами, существующими в языке. Так, С. Н. Лебедев, анализируя функционирование концептов «менеджмент» и «управление» в русской культуре, утверждает, что термин «управление» гораздо более объемный и содержательный, имеет более широкую сочетаемость, может относиться к разным системам (биологическим, техническим, социальным), в то время как заимствование «менеджмент» тесно связано с признаком «научный» и профессиональной деятельностью в области управления организациями [3]. Еще одной особенностью внедрения термина в принимающую

лингвокультуру становится общая для всех терминов, существующих в языке, тенденция к его постепенной детерминологизации вследствие активного употребления в текстах массмедиа: «В общественном сознании существуют традиции освоения терминов, которые имеют повышенную значимость в плане выражения важных для этноса или социума идей или ценностей в тот или иной исторический период. Культурными механизмами внедрения научных понятий в массовый обиход является их повсеместное нетерминированное употребление в языке политических и общественных деятелей, в медийном дискурсе, откуда они проникают и в обыденную речевую практику носителей языка» [4, с. 56].

Проведенный анализ употребления заимствования «инвестиция» в газетном подкорпусе Национального корпуса русского языка показал, что первая его фиксация в корпусе приходится на 1983 год, где оно обозначает термин экономического дискурса и относится в основном к описанию экономической ситуации разных стран: *«французский патронат – объединение предпринимателей, представляющих интересы крупного капитала, – отказывается делать инвестиции в промышленность»* (Капитализм без прикрас. Уровень жизни или уровень нищеты? // Аргументы и факты, 1983.11.15) [5]. В целом в этот период термин «инвестиция» в русской лингвокультуре приобретает негативный оттенок значения, связанный с восприятием им политико-идеологической окраски в рамках противостояния социалистической и западной экономических систем, о чем можно понять из заголовков статей, где используется заимствование «инвестиция»: «Капитализм против человека», «Разоблачение: Чили», «Швеция: проблемы экономики» и т. д. В примерах показаны экономические махинации буржуазии как классового врага процветания: *«Иностранные инвестиции оседали в сейфах ТНК и в карманах национальной буржуазии»* (Разоблачение. Чили: фальшивый расцвет // Аргументы и факты, 1983.03.15) [5]. При этом остро стоит вопрос «мирных» инвестиций как залога экономической и политической стабильности: *«По существующим подсчетам, инвестиции в мирные отрасли экономики могут создать больше рабочих мест, чем соответствующие вложения в военное производство»* («Нет» войне! Предотвратить ядерную катастрофу // Аргументы и факты, 1983.05.31) [5].

Активно использоваться заимствование «инвестиция» начало с середины 90-х годов XX века, когда стали происходить сдвиги в экономической жизни страны. В это время произошло заимствование многочисленных терминов экономического дискурса, среди которых основополагающим, выражающим новые идеи хозяйствования стал «маркетинг»: «Маркетинг был институализирован, приняты его идеи – продавать то, что потребляется, а не потреблять то, что производится» [6, с. 48]. Отметим, что в это время термин «инвестиция» употребляется в основном с эпитетом «иностранный» и приобретает негативную окраску: *«Очень важное поле нашей деятельности – приватизация предприятий, содействие осуществлению иностранных инвестиций в Россию, имущественные споры о недвижимости»* (Предпринима-

телю на заметку. Без юриста – никуда! // Аргументы и факты, 1992.09.09) [5], что отражает социально-экономическую ситуацию в стране того времени, а также трудности привыкания к новым экономическим реалиям. В настоящее время в результате эволюции и когнитивно-языкового регулирования заимствование «инвестиция» употребляется в различных сферах, приобретая множество метафорических репрезентаций, что указывает на присвоение русской культурой данного заимствования. Рассмотрим особенности концептуализации заимствования «инвестиция» в текстах политической сферы, по количеству употреблений стоящей на втором месте после экономической сферы (8449 текстов, 12 170 примеров).

Метафоры в текстах политической сферы направлены в векторном направлении «друг-враг», выражая стратегию соперничества или, наоборот, стратегию сотрудничества. С позиции сотрудничества инвестиция может приобретать положительное значение, связанное с восприятием ее как помощи, направленной на развитие чего-либо, используется со словами «мир», «поддержка», «благополучие», «безопасность», создающими позитивный эмоциональный фон. Инвестиции становятся средством установления взаимовыгодных отношений, способствующих процветанию чего-либо: «сотрудничество в сфере инвестиций и торговли».

Слово «инвестиция» в отдельных контекстах используется в качестве положительно оцениваемого эвфемизма для объяснения собственных действий как приносящих благо (направленность на сотрудничество). При разворачивании в СМИ другой стороны контексты приобретают ироничную тональность, поскольку слово «инвестиция» воспринимается как используемое для маскировки политических мотивов, прикрывающих под видом помощи заботу о собственных интересах (направленность на борьбу).

Заимствование «инвестиция» перестает употребляться в своем прямом терминологическом значении для обозначения выгодного вложения денег, которые приносят прибыль инвестору и помогают в подъеме экономики в целом. Заимствование «инвестиция» относится к денежным вложениям, а также к затратам времени, сил и т. д. и касается получения не только финансовой выгоды, но и выгоды в других стратегически важных аспектах. Так, в политической сфере при использовании в ситуации соперничества, заимствование «инвестиция» связано с выражением позиции лидерства, подавлением конкурента, например, «инвестиции в «стратегическое соперничество» с ними», «инвестиции в технологическое превосходство».

Сочетаемость концептуального заимствования «инвестиция» в текстах политической сферы показывает, что оно начинает рассматриваться в качестве орудия политической борьбы, что зафиксировано в метафорических выражениях «наращивать инвестиции» (сравним: наращивать вооружение); «усиливать инвестиции» (в ряду однородных членов со словами «позиции, возможности, планы, друзей и союзников») Эти выражения переводят нас из чисто экономической плоскости в плоскость соперничества за сферу политического влияния, когда инструментом экономической войны служит

инвестиция. Еще одна сочетаемость, связанная с метафорами борьбы, «защищать инвестиции». Для инвестиций что-то может представлять угрозу, например, противодействие развитию перспективных планов или идей: *«Пекин будет защищать инвестиции в осуществляемый им проект»* (Пентагон: Китай будет строить новые зарубежные военные базы // Московский комсомолец, 03.05.2019) [5].

«Инвестиция» начинает употребляться в контексте разговорной лексики, например, в метафорических оборотах «инвестиции пойдут насмарку», «инвестиции утекают в никуда». Она используется в тексте СМИ, чтобы придать речи выразительность, выражая эмоции сторон: *«И теперь сделанные в них «инвестиции» «вот-вот пойдут насмарку»»* (Эрдогана обвинили в намеренной атаке на Су-24 для решения собственных политических задач // NEWSru.com, 25.11.2015) [5]. В данном контексте слово «инвестиция» берется в кавычки, указывая, что, на самом деле, преследуется только цель расширения сферы своего влияния, а не реальная забота о благополучии людей.

Ключевым в восприятии заимствования «инвестиция» в политической сфере становится слово «выгода». Инвестиция становится инструментом воздействия, позволяющим оказывать давление и добиваться своего, появляется выражение «отмывать инвестиции».

Метафоры часто переносят политические отношения в сферу межличностных. Инвестиции могут «угрожать», «душить», «обхаживать», «отпугнуть» — очевидно, что употребление данных глаголов указывает на различные средства манипулирования, которые способствуют усилению политического влияния. Так, использование слова «обхаживать» (объектная сочетаемость) переводит нас в плоскость человеческих отношений, косвенно указывая на их манипулятивный характер, представляя действия не как объективную необходимость, но как целенаправленное воздействие с целью получить популярность и политическую поддержку («обхаживать» — заискивать во всем угождать, добиваясь чего-либо).

Метафоризация происходит при адъективной сочетаемости заимствования «инвестиция», отражающей его переосмысление в контексте связи с личными переживаниями: словосочетание «тревожные» инвестиции оценивается не с позиции риска потери прибыли со стороны инвестора, а как угроза политического и военного усиления. Еще один пример несвойственной для данного слова сочетаемости — «назревшие инвестиции» («назревший» — требующий решения, насущный: «назревшее решение проблемы»), т. е. в данном контексте подразумеваются давно уже необходимые и требующиеся капиталовложения.

Сделаем вывод, что заимствование «инвестиция» в русской культуре в процессе эволюции прошло процесс детерминологизации, употребляясь в современной лингвокультуре в различных сферах жизни, приобрело переносные значения, обозначая в политической сфере не только экономически выгодное вложение денег, но и действия, направленные на усиление полити-

ческого влияния. Проанализированные контексты показывают, что заимствование «инвестиция» стало средством манипулирования общественным сознанием, создания образа друга или врага, в соответствии с которым выбираются стратегии сотрудничества или соперничества, определяющие использование языковых репрезентаций. Так, метафорические контексты переносят политические отношения в сферу межличностных отношений, делая читателей вовлеченными и более заинтересованными. Экономический термин приобретает личностную окрашенность через ассоциирование с человеческими эмоциями, связанными с оценкой значения тех или иных инвестиций. Используются метафоры военной сферы, показывающие происходящую политическую борьбу и экономическую конкуренцию. Кроме того, происходит манипулирование на основе восприятия инвестиций как способствующих развитию и процветанию, когда заимствование «инвестиция» превращается в эвфемизм для обозначения действий, направленных на получение собственной выгоды. Таким образом, когнитивно-языковое регулирование использования заимствования «инвестиция» в русской лингвокультуре позволило ему приобрести уникальные черты, отражающие национальную картину мира, и определило восприятие и оценку данного заимствования в процессе развития.

ЛИТЕРАТУРА

1. Ефимова А. Д. Лингвокультурная специфика медиатизации заимствованных концептов в СМИ // Медиатизация лингвокультурных феноменов: аспекты языковой репрезентации: монография / Е. Ю. Ильинова, А. Д. Ефимова, Л. А. Кочетова, И. В. Кононова, В. А. Митягина, И. В. Палашевская, Л. Н. Ребрина, Е. В. Трощенко [и др.] ; под общ. ред. д-ра филол. наук, проф. Л. А. Кочетовой ; Федер. гос. авт. образоват. учреждение высш. образования «Волгогр. гос. ун-т». Волгоград: Изд-во ВолГУ, 2024. 220 с.
2. Карасик В. И. Языковое проявление личности. М.: Гнозис. 2015. 382 с.
3. Лебедев С. Н. «Управление», «Менеджмент», «Администрирование»: проблема концептов // Вестник Российского университета дружбы народов. Серия Социология. № 4. 2011. С. 143–151.
4. Жуковская Л. И. Семантическое наполнение концепта «менталитет / ментальность» и его языковое воплощение в современном русском языке: дис. ... канд. филол. наук. Нижний Новгород. 2015. 197 с.
5. Национальный корпус русского языка [Электронный ресурс]. URL: <https://ruscorpora.ru> (дата обращения: 15.05.2025).
6. Естественные языки на автохтонных территориях и за их пределами: коллективная монография. Москва, 2024. 223 с.

УДК 81-114.4

Ибадова Нигяр Салех кызы

педагог

Бакинский Славянский Университет

г. Баку, Азербайджан

Ibadova Nigar Saleh

Lecturer

Baku Slavic University

Baku, Azerbaijan

ibadowa.nigar@gmail.com

КОРПУСНАЯ ЛИНГВИСТИКА: СОВРЕМЕННЫЙ ПОДХОД К ИЗУЧЕНИЮ ЯЗЫКА

В статье рассматриваются теоретические и практические аспекты корпусной лингвистики – современного направления в языкознании, основанного на систематическом сборе и анализе больших массивов текстов. Особое внимание уделяется характеристикам корпусов, функциям метаданных и их значению для научных исследований. Рассматривается опыт Азербайджана в развитии корпусной лингвистики в контексте глобализации и цифровизации языка.

Ключевые слова: корпусная лингвистика; лингвистический корпус; аннотация; метаданные; частотный анализ; азербайджанский язык.

CORPUS LINGUISTICS: A MODERN APPROACH TO LANGUAGE STUDY

This article examines the theoretical and practical aspects of corpus linguistics – a contemporary field of linguistic research based on the systematic collection and analysis of large text corpora. Special attention is paid to the characteristics of corpora, the functions of metadata, and their significance for scientific research. The article also explores Azerbaijan's experience in developing corpus linguistics in the context of globalization and digitalization of language.

Key words: corpus linguistics; linguistic corpus; annotation; metadata; frequency analysis; Azerbaijani language.

Corpus linguistics is a field of linguistics based on the systematic gathering, storage, and analysis of extensive text corpora using computer technologies. It provides researchers with a unique opportunity to study language based on real data rather than only intuitive judgments or isolated examples. Today, corpus linguistics is one of the fastest-growing fields in linguistic science.

What is a Corpus? A corpus is an electronic database of texts collected for scientific analysis. The texts may include fiction, scientific articles, newspaper publications, spoken speech, subtitles, and even social media content. Corpora are usually annotated – grammatical, semantic, and other tags are added to the words, facilitating analysis. According to Lemnitzer and Zinsmeister, a corpus is a collection of written or spoken utterances, typically stored in a machine-readable digital format.

Each text in a corpus includes not only the linguistic data itself but may also contain metadata (information about origin, genre, author, etc.) and linguistic annotation that describes grammatical, semantic, and other features of the text.

Why is Corpus Linguistics Important? Traditionally, linguistics has relied on descriptive theories and the intuition of the researcher. Corpus linguistics allows us to:

- Obtain objective data on the frequency and distribution of words and constructions;
- Study actual linguistic trends and changes;
- Analyze word and phrase usage in context;
- Create dictionaries and reference works based on real language use;
- Support language teaching and machine translation.

Lemnitzer describes corpus linguistics as a scientific discipline that deals with the description of natural language utterances and their elements and structures. These descriptions are based on the analysis of authentic (real) texts collected in corpora, and linguistic theories are derived from them. Corpus linguistics, as a scientific field, must adhere to scientific principles and standards.

Language descriptions based on corpora are used in various fields, such as language teaching, documenting endangered languages, dictionary development, and natural language processing.

Corpus Collection and Construction A corpus is a carefully assembled collection of texts or speech data. Researchers determine the corpus's goal and scope: for example, it may consist solely of scientific articles, interviews, or social media texts.

The preparation steps include:

1. **Data cleaning** – removing errors, duplicates, or non-informative elements (ads, code);
2. **Standardization** – converting texts into a unified format (encoding, fonts);
3. **Annotation** – tagging words with parts of speech (POS tags), lemmas, grammatical features.

According to Lemnitzer, linguistic corpora usually have the following characteristics:

- Representativeness;
- Metadata support;
- Linguistic annotation.

The first feature – representativeness – distinguishes corpora from other linguistic data collections. While large corpora typically have all the listed features, smaller research corpora may not include all of them.

It is particularly important to prioritize oral texts, as spoken language is the primary medium of human communication.

These criteria cover both extralinguistic and pragmatic-semantic aspects. For system-oriented linguists, it is essential that structural variation is represented in

the corpus. In well-documented languages, completeness can be ensured through targeted selection. In under-documented languages, this is impossible because the desired structures are not yet known. However, it is assumed that structural variation will naturally emerge if functional variation is adequately covered – following the principle “form follows function”.

Modern corpora usually consist of full texts or dialogues – for example, newspaper articles or chat transcripts. Text length varies widely: from short messages like SMS to long texts such as novels. For many studies, wide textual contexts are required to analyze phenomena like pronoun reference or discourse structure.

The Role of Metadata Metadata serve several key purposes:

- Record contextual aspects of text creation (e.g., time, place, author);
- Allow filtering and categorization for research.

Lemnitzer provides the example: “If the films in an archive are annotated with appropriate metadata, you can find all movies where Woody Allen was involved only as a director.”

He also notes two levels of metadata:

1. For the digital file itself (e.g., its name, format);
2. For the original source material (e.g., a newspaper article, diary page).

Metadata enable targeted classification, both manual and automated search, and analysis in corpora.

Quality in Corpus Compilation

Text quality involves phonetic, grammatical, and stylistic accuracy as well as meaningful content. Documentation is created for posterity, so substandard texts are neither useful nor resource-worthy. Quality also includes the compiler’s work: recording quality, transcription accuracy, annotation, and metadata reliability. Quality must take precedence over quantity.

Some corpora are based on transcriptions of medieval manuscripts in different dialects and scripts. Editions may include:

- Adapted transcription close to modern norms;
- Diplomatic transcription preserving original form with minor clarifications;
- Facsimile reproductions close to the original layout.

Corpus Linguistics Methods Documenting language varieties and genres is not new. However, documenting an entire language arose in the 1980s with the realization that many languages were endangered. Full-scale documentation became a priority to preserve linguistic heritage.

Specialized tools – corpus managers and search systems – allow researchers to:

- Search word usage cases;
- Analyze collocations;
- Study grammatical structures.

Analytical methods include:

- **Frequency analysis** – how often a word or structure appears;
- **Concordance** – viewing all contexts of a word;
- **Statistical analysis** – identifying patterns and correlations.

Corpus searches often utilize morphological descriptors, which consider grammatical categories and find all relevant forms. Searches can span the entire corpus or specific context-defined subsets. This is useful in empirical material collection, dictionary development, and grammatical analysis.

Examples of Corpora:

- British National Corpus (BNC);
- Russian National Corpus (RNC);
- Corpus of Contemporary American English (COCA);
- Specialized spoken or genre-based corpora.

Data reflect phenomena from scientific domains and are accepted as evidence based on methodological rigor. In scientific reasoning, data substitute the actual phenomenon, often inaccessible directly. In inductive approaches, data serve as evidence; in deductive methods, they test hypotheses.

For centuries, traditional grammar preserved syntactic and morphological rules. Corpus-based research often yields surprising results about actual language use. Corpus linguistics is just one of many methods – it complements, but does not replace, elicitation techniques like minimal pairs. As corpora grow, their empirical value increases.

If a form does not appear in a speaker's lifetime exposure, it is inferred by analogy – the same applies to linguists using finite corpora. Larger corpora increase the reliability of such inferences.

Corpus Linguistics in Azerbaijan

In Azerbaijan, corpus linguistics was introduced through a state program to promote the national language in the context of globalization. The program included the creation of various digital dictionaries – orthographic, explanatory, terminological, and frequency-based.

Building a national corpus is a strategic state priority. It helps preserve the language, monitor lexical changes over time, and facilitates research without manual data collection. Researchers can extract and analyze linguistic units in natural contexts.

Compiling frequency dictionaries and concordances for classic Azerbaijani authors – Nizami Ganjavi, Mirza Alakbar Sabir, and Muhammad Fuzuli – is a key step toward creating individual corpora for these poets. These projects serve as a foundation for the Azerbaijani National Corpus.

A comprehensive national corpus must represent all functional styles and genres, including classical literature, folklore, and oral traditions. Frequency dictionaries are also critical for machine translation, summarization, authorship attribution, publishing, and forensic linguistics.

Conclusion

Corpus linguistics is a powerful tool for modern language research, enabling deep, objective insights into language structure and use. As technology and data availability grow, corpus methods gain increasing relevance in research, education, translation, and linguistic technology development.

REFERENCES

1. Lemnitzer, L., & Zinsmeister, H. (2010). *Korpuslinguistik: Eine Einführung*. Tübingen: Narr Francke Attempto Verlag.
2. Rayskina, V. A., & Dubnyakova, O. A. (n.d.). *Современные методы корпусной лингвистики при анализе текста (на примере корпуса BFM)*. Retrieved from <https://cyberleninka.ru/article/n/sovremennye-metody-korpusnoy-lingvistiki-pri-analize-teksta-na-primere-korpusa-bfm/viewer>
3. Lehmann, C. (n.d.). *Daten–Korpora–Dokumentation*. Retrieved from <https://www.christianlehmann.eu/publ/daten.pdf>
4. Khalilova, G. A. (n.d.). *Развитие азербайджанской корпусной лингвистики. Создание корпуса текстов азербайджанских авторов (поэтов и писателей) и национального корпуса азербайджанского языка*. Retrieved from <file:///C:/Users/Администратор/Downloads/razvitie-azerbaydzhanskoy-korpusnoy-lingvistiki-sozdanie-korpusa-tekstov-azerbaydzhanskih-avtorov-poetov-i-pisateley-i-natsionalnogo-korpusa-azerbaydzhanskogo-yazyka.pdf>
5. Azerbaijani National Corpus. (n.d.). Retrieved from <https://korpus.azerbaycandili.a>

УДК 81-139

Кирина Маргарита Александровна
младший научный сотрудник Лаборатории
языковой конвергенции
Национальный исследовательский
университет «Высшая школа
экономики»
г. Санкт-Петербург, Россия

Margarita Kirina
Junior Research Fellow of the Linguistic
Convergence Laboratory
National Research University «Higher
School of Economics»
St. Petersburg, Russia
mkirina@hse.ru

Лукьянчикова Алиса Сергеевна
стажер-исследователь Лаборатории
языковой конвергенции
Национальный исследовательский
университет «Высшая школа
экономики»
г. Санкт-Петербург, Россия

Alisa Lukyanchikova
Research Assistant of the Linguistic
Convergence Laboratory
National Research University «Higher
School of Economics»
St. Petersburg, Russia
lukyanchikova@edu.hse.ru

«ЗВЕЗДЫ РЕКОМЕНДУЮТ ВЕСАМ ПИТЬ СЛИВОВОЕ ВИНО»: ИССЛЕДОВАНИЕ АСТРОЛОГИЧЕСКОГО ДИСКУРСА НА ОСНОВЕ РАСПРЕДЕЛЕНИЙ ЧАСТОТНОЙ ЛЕКСИКИ И СЕНТИМЕНТ-АНАЛИЗА

В статье рассматриваются характерные особенности гороскопических текстов как части астрологического дискурса. Материалом исследования выступает представительная выборка ежедневных предсказаний на русском языке, опубликованных в открытых группах социальной сети «ВКонтакте», суммарным объемом 1185425 словоупотреблений. С использованием методов корпусной и компьютерной лингвистики анализируются содержательные лексические единицы – как общие, так и отличительные для каждого знака зодиака (в сопоставлении с устоявшимися архетипами). Кроме того, проводится сравнение верхней зоны частотных словарей, полученных на основе собранного корпуса гороскопов и подкорпуса «гороскопы», входящего в Национальный корпус русского языка (НКРЯ). Немаловажным этапом настоящей работы стало определение эмоциональной валентности текстов в соответствии с тринарной классификацией (положительная, отрицательная, нейтральная). Анализ тональности, проведенный с использованием словаря оценочной лексики RuSentLex, позволил установить, что гороскопы имеют преимущественно положительную или нейтральную эмоциональную окраску, а в среднем наиболее позитивным оказался поджанр «любовный гороскоп». В результате исследования с опорой на количественные данные удалось выявить специфические черты современных астрологических предсказаний на лексическом уровне и в контексте их эмоциональной «заряженности», а также отметить некоторые ограничения изучения этого вида текстов.

Ключевые слова: астрологический прогноз; корпус гороскопов; зодиакальные архетипы; корпусная лингвистика; частотный словарь; анализ тональности.

“STARS RECOMMEND LIBRAS TO DRINK PLUM WINE”: A STUDY OF ASTROLOGICAL DISCOURSE BASED ON WORD FREQUENCY DISTRIBUTIONS AND SENTIMENT ANALYSIS

This article explores the distinctive features of horoscope texts as part of astrological discourse. The research material consists of a representative sample of daily horoscopes in Russian, published in open groups on the social network VKontakte, with a total volume of 1,185,425 word occurrences. Using methods from corpus and computational linguistics, the study analyses substantive lexical units—both general and unique to each zodiac sign (in comparison with established archetypes). Additionally, a comparison is conducted between the most frequent words derived from the collected horoscope corpus and those from the “horoscopes” subcorpus of the Russian National Corpus (RNC). An important stage of the work involved determining the emotional valence of the texts according to a trinary classification (positive, negative, neutral). Sentiment analysis based on the RuSentiLex evaluative lexicon revealed that horoscopes predominantly have a positive or neutral emotional tone, with the “love horoscope” subgenre being on average the most positive. As a result, relying on the quantitative data, it was possible to identify lexical and emotional characteristics of modern astrological forecasts, as well as to note some limitations when studying this type of text.

Key words: astrological forecast; horoscope corpus; zodiac archetypes; corpus linguistics; frequency dictionary; sentiment analysis.

Астрологический прогноз – особый жанр, получивший широкое распространение в медийном пространстве постсоветской действительности, который берет свое начало еще в донаучном периоде развития общества. Понятие гороскопа, не имея четких границ и исключительного определения, сегодня во многих культурах воспринимается как предсказание на день. Хотя сфера бытования гороскопов крайне обширна, из-за их маргинального, даже псевдонаучного, статуса [1], эти тексты не столь популярны в научном сообществе. Тем не менее такие особенности гороскопического текста, как функциональная направленность, рекомендательный характер, а также специфика выражения модальности и оценки, вызывают лингвистический интерес и описываются в ряде исследований [2–3]. Языковые средства, свойственные этим прогнозам, рассматриваются в рамках как магического [4], так и астрологического дискурсов [5]. Неслучайно характерные признаки названных текстов изучаются и как черты медиажанров [6]: гороскопы встречаются в журналах, газетах, распространяются по радио, публикуются на форумах в сети Интернет.

Популярность астрологических прогнозов среди широкой аудитории объясняется конструируемой «персонализированностью» текстов – каждый с нетерпением ждет, что сегодня скажут «про него». Более того, считается, что за знаками зодиака закреплены определенные архетипы [7], обобщение которых представлено, в соответствии с [2; 8], в Таблице 1. Например, Овна относят к образу воина, Тельца – к образу хранителя [Там же]. В этой связи интерес представляет то, в какой степени гороскопическое предсказание оказывается в действительности «адресным», то есть насколько в нем (намеренно или нет) отражаются ассоциируемые с тем или иным знаком зодиака стереотипы.

Т а б л и ц а 1

Характеристика знаков зодиака по архетипам

№	Знак зодиака	Архетип
1	Овен	Воин, первооткрыватель
2	Телец	Хранитель
3	Близнецы	Коммуникатор, посредник
4	Рак	Хранитель домашнего очага (мать)
5	Лев	Король, созидатель
6	Дева	Служитель
7	Весы	Миротворец
8	Скорпион	Преобразователь
9	Стрелец	Философ, учитель
10	Козерог	Стратег
11	Водолей	Новатор, гуманист
12	Рыбы	Мечтатель, мистик

Цель настоящего исследования – проследить особенности лексического состава гороскопов, в том числе проверить, проявляются ли архетипические черты в астрологических прогнозах. Под гороскопами будут пониматься ежедневные предсказания, то есть тексты, которые предписывают то или иное поведение или пророчат некоторые события в определенный день согласно знаку зодиака человека. Методологической основой этой работы станет корпусный подход. В отличие от классических методов, компьютерные инструменты позволяют, во-первых, принять во внимание больший объем данных, а во-вторых – независимо от интроспективных допущений, рассмотреть общие лексико-тематические особенности групп текстов, разделенных в нашем случае в зависимости от целевой аудитории, то есть соответствующего знака зодиака.

Материалом выступили гороскопы, размещенные в открытых группах «ВКонтакте», которые занимаются регулярной публикацией ежедневных астрологических прогнозов для каждого из 12 знаков зодиака с 2014 по 2025 гг. (включительно)¹. Пример выгрузки, за исключением некоторых не используемых в рамках этого исследования метаданных (количество лайков, комментариев, репостов, просмотров, оставленных пользователями), представлен в Таблице 2. В «сыром» виде полученный датасет насчитывал 90 642 строки (около 17 тыс. строк на каждый знак зодиака). Далее осу-

¹ Вопрос о возможности попадания в корпус текстов, сгенерированных искусственным интеллектом, остается за рамками текущего исследования.

ществлялась очистка данных в несколько этапов. Так, были удалены тексты, не являющиеся гороскопами вида «ежедневное предсказание»: рекламные и поздравительные посты, гороскопы на неделю, год и иные временные промежутки, а также гороскопы, составленные не в соответствии с зодиакальной системой (например, кельтские гороскопы). Затем автоматически были «развернуты» посты, которые в одном тексте включали несколько видов, – общий, бизнес- и любовный гороскопы.

Т а б л и ц а 2

Пример выгрузки (фрагмент датасета)

ID	Заголовок	Текст	Год публикации	Знак зодиака
1436	Любовный гороскоп на 16 марта 2025 года	Спокойствие, только спокойствие – так проще достичь желаемого!	2025	Стрелец
4911	Гороскоп на 30 мая 2024 года	Нынче у вас должна обостриться интуиция. Ей и следуйте.	2024	Лев
31540	Гороскоп на 25 апреля 2018	Наиболее трудные дела пока отложите.	2018	Рак
33931	Бизнес-гороскоп на 01 октября 2017	Будьте осторожны в финансовых делах. Возможны происки неожиданных конкурентов.	2017	Телец
46142	Общий гороскоп на завтра, 9 сентября, 2016	Вы прекрасный исполнитель и способны многого добиться. Сегодня Вам предстоит работать с общественностью.	2016	Козерог

После удаления «мусорных» постов и гороскопов-дублей (без учета знака зодиака), токенизации и лемматизации, выполненных с помощью MyStem [9], а также сегментации на предложения с использованием библиотеки NLTK [10], объем корпуса составил 28 149 строк или 1 185 425 словоупотреблений (15578 лексических типов). Средний размер группы гороскопов, приходящихся на каждый знак зодиака, равен 2346 текстам со средним суммарным объемом 99 427,42 токена. Текст гороскопа как единица анализа имеет следующее лингвостатистическое описание (в средних значениях): длина в словах – 42,39 (σ – 13,97), длина в предложениях – 3,42 (σ – 1,18).

Перейдем к результатам. Сначала рассмотрим лексические особенности гороскопических текстов. Верхняя зона частотного словаря, составленного по данным всего корпуса, представлена в Таблице 3, а характерные содержательные лексемы на Рисунке 1. На примере наиболее релевантных сочетаний слов прослеживаются прагматико-дискурсивные черты подобного текста: предсказание событий на завтра (*день, завтра*), ориентация на положитель-

ный исход дня (*благоприятный, день*), а также межличностные отношения и рекомендации по их организации, в частности через открытость к новым встречам и знакомствам (*человек, новый, знакомство, отношение*).



Рис. 1. Сетевая визуализация совстречаемости наиболее частотных слов (исключая стоп-слова)

Стоит отметить, что самые распространенные слова, то есть такие, которые встречаются в каждом документе корпуса, составляют 18,83 % от общего числа лексических типов (с учетом повторений – 92,27 % от объема всего корпуса). При этом уникальные для каждого из подкорпусов (то есть знаков зодиака) единицы не оказываются интерпретируемыми и в своем большинстве не обнаруживают статистически значимой разницы в частотных распределениях (эксперимент по извлечению ключевых слов осуществлялся при использовании Log-Likelihood и пороговом значении $p\text{-value} < 0,05$, в качестве референтного для каждого целевого корпуса брались оставшиеся 11). Однако некоторые наблюдения путем фронтального анализа лексико-тематических групп и характерных архетипических мотивов представляется возможным. Обратимся для этого к Рисунку 2, на котором для каждой «зодиакальной» выборки визуализированы по 50 наиболее частотных содержательных и не входящих во множество общих для всего корпуса лексем.

Например, в облаке слов, характеризующем Овна заметен фокус на карьеру (*шеф, торговля, совещание*) и советы «заботиться» (возможно, из-за стереотипов недостатка созидательности у представителей этого знака зодиака). Значительное место занимает лексика, связанная со знаниями, их получением и применением (*образование, изучение, философский*). В отношении Тельца заметна «предостерегающая» лексика (*угроза, агрессивный*), а также связанная с работой – примечательно, что появляется фигура подчиненного (*работник, офис, совещание*), а также материальный аспект (*вложение*). Вновь появляется указание на сферу взаимоотношений (*привязанность, заботиться*). Тема работы наблюдается и в частотных распределениях гороско-

[illegible]

79

Любопытно, что при всей доминантной для архетипа созидательности (*возобновлять, благосклонно*), в подкорпусе Рака мало представлена рабочая сфера. Однако отмечаются слова, описывающие отношение к жизни (*практичность, трудолюбие, знание*) и к людям (*привязанность, верность*), а также экспрессивная, говорящая о чувственном восприятии действительности лексика (*всевозможный, потрясать, головокружительный, вспыхивать, напряженность*). Для Скорпиона выделяются лексико-тематические группы со значением богатства/успеха (*щедрость, благосостояние, состояние, успех, вложение, зарплата*), а также «напряжение – отдых» (*повеселиться, напряженность*). Однако это в большей мере отражает тематическую направленность гороскопов, предсказания в которых строятся с опорой на сферу рабочих или личных отношений, чем указывает на особенности характера этого знака зодиака. Дева же, напротив, описывается через слова, связанные с логикой, аналитическим мышлением (*убедительный, анализ, довод, построение*), что согласуется со стереотипом о ее педантичности.

Несмотря на отмечаемые в некоторых случаях различия, проведенный анализ указывает на необходимость обращения к гороскопу как к единому дискурсивному полю, отличающемуся слабо выраженной дистинктивностью (по крайней мере, если в качестве дифференцирующего параметра рассматривать знак зодиака). В связи с этим мы провели еще один эксперимент, а именно сравнили частотные распределения лемм в пользовательском корпусе, описанном в этом исследовании (и условно названном «гороскопы в соцсетях»), и в подкорпусе гороскопов Национального корпуса русского языка (НКРЯ) [11]. Подкорпус «гороскопы» НКРЯ насчитывает 130 текстов суммарным объемом 105877 словоупотреблений. Как позволяют судить метаданные, гороскопы, включенные в НКРЯ, были составлены в период с 1991 по 2010 гг. (включительно) и представляют собой принципиально другой вид медиа. Так, их источниками выступают журналы, газеты, а также, что немаловажно, встречаются авторские астрологические прогнозы.

В результате сравнения верхней зоны частотных словарей (см. Таблицу 3) стала заметна ориентированность гороскопов, публикуемых в социальных сетях, на прогнозирование событий завтрашнего дня, в то время как гороскопы в журналах и газетах, что несколько ожидаемо, охватывали конкретный *день* или *неделю* целиком. Объяснение находится в особенностях сферы функционирования этих текстов, а также в исключении из нашего корпуса не ежедневных предсказаний. Кроме того, кажется, что предсказания из соцсетей стремятся быть более личными (см. ранг местоимения *вы*), отсутствие чего в подкорпусе гороскопов из НКРЯ, вероятно, является метой ограничений, накладываемых на периодические издания, их внешней модулируемостью.

В обоих корпусах заметна тема рабочих отношений. Однако в подкорпусе НКРЯ лемма *работа* имеет ранг 24, а в корпусе «гороскопы в соцсетях» – 50. Похожая тенденция на «опущение» по рангу заметна и для

некоторых других лексем: *здоровье* – 31 vs. 234, *любовь* – 34 vs. 87, *деньги* – 37 vs. 289. По всей видимости, для ежедневных онлайн-предсказаний, хоть и важна, но в меньшей мере свойственна привязка к какому-либо конкретному аспекту жизни, в результате чего они представляются более общими и несколько абстрактными. В то время как для публикуемых в официальных СМИ гороскопах характерна выраженная категоризация «по сфере действия».

Т а б л и ц а 3

Сравнение частотных словарей:
гороскопы в соцсетях vs. подкорпус НКРЯ (верхняя зона)

Гороскопы в соцсетях				НКРЯ (подкорпус "гороскопы")			
Ранг	Лемма	Частота		Ранг	Лемма	Частота	
		Абс.	IPM			Абс.	IPM
1	и	40706	34338,74	1	и	3954	37345,22
2	вы	38437	32424,66	2	в	3297	31139,91
3	в	32133	27106,73	3	не	1831	17293,65
4	не	29601	24970,79	4	с	1675	15820,24
5	быть	22860	19284,22	5	на	1615	15253,55
6	на	17127	14447,98	6	вы	1612	15225,21
7	что	15461	13042,58	7	быть	1338	12637,31
8	с	14595	12312,04	8	что	1024	9671,60
9	то	14137	11925,68	9	а	862	8141,52
10	ваш	13956	11772,99	10	по	768	7253,70
11	день	12758	10762,38	11	к	675	6375,32
12	мочь	11737	9901,09	12	для	655	6186,42
13	свой	11717	9884,219	13	это	617	5827,52
14	сегодня	10877	9175,612	14	то	613	5789,74
15	человек	10506	8862,644	15	свой	581	5487,50
16	для	10174	8582,58	16	мочь	569	5374,16
17	он	8644	7291,90	17	но	556	5251,38
18	или	8568	7227,79	18	ваш	555	5241,93
19	к	8370	7060,76	19	он	518	4892,47
20	это	8214	6929,16	20	этот	511	4826,36

21	если	7995	6744,42	21	день	510	4816,91
22	этот	7722	6514,12	22	если	507	4788,58
23	завтра	6574	5545,69	23	неделя	467	4410,78
24	они	6209	5237,78	24	работа	459	4335,22
25	который	6032	5088,47	25	или	451	4259,66

Следующим шагом этой работы стало выявление особенностей эмоционального воздействия гороскопических текстов, определяемых на лексическом уровне. Для этого с использованием русскоязычного словаря оценочной лексики RuSentiLex [12] и программного обеспечения Orange [13] был проведен анализ тональности (или сентимент-анализ). Задача анализа тональности сводится к определению текста, выражающего мнение, и его характеристики с точки зрения направленности и степени выраженности эмоционального отношения (положительного, отрицательного или нейтрального) [14]. Расчет оценки тональности в настоящем исследовании производился в соответствии с [15] по следующей формуле:

$$(sum(pos) - sum(neg)) / document\_length * 100,$$

где pos – количество положительных слов (в типах), neg – количество отрицательных слов (в типах), а document_length – длина документа (количество слов в тексте) [16].

В результате были получены следующие распределения по количеству текстов: положительные – 12093 (42,96 %), нейтральные – 10788 (38,32 %), негативные – 5268 (18,71 %), что согласуется с заключениями, представленными в [17]. Так, и в нашем случае, кажется, что лексико-эмоциональное наполнение текстов имеет целью вызвать у читателя положительные эмоции: спокойствие, надежду, стабильность. Это гарантируется чтением предсказания, что функционально выступает действием эквивалентным возможности контролировать свое будущее. При этом, как видно исходя из Рисунка 3, знак зодиака не влияет на эмоциональную направленность текста, что вновь свидетельствует о генерализованном характере астрологических предсказаний.

Приведем несколько примеров положительных (1–2), негативных (3) и нейтральных предсказаний (4–5):

(1) *Сегодня вы одержите пусть небольшую, но победу. Этот день принесет вам уважение окружающих.*

(2) *Сегодня вы просто обречены на успех в амурных делах. Не играйте в карты!*

(3) *Сегодня вы рискуете допустить ошибку или в чем-то просчитаться. Вам могут предоставить недостоверную информацию, так что будьте бдительны.*

(4) *Сегодня лучше действовать сообща. Одному столько не выпить.*

(5) *На работе не будет происходить ничего особенного. Будете выполнять рутинные обязанности и мелкие поручения.*

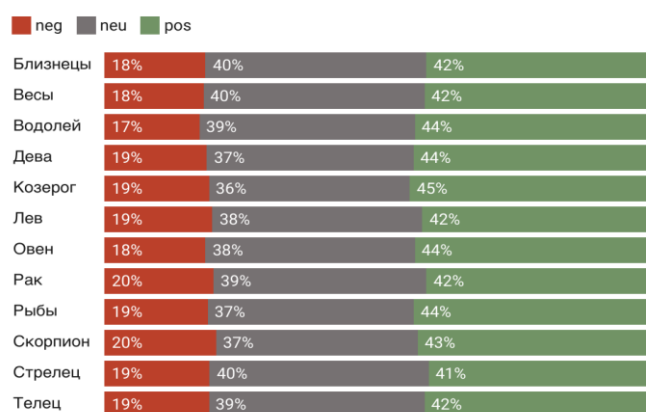


Рис. 3. Распределение гороскопов, сгруппированных в зависимости от подкорпуса, по тональности (в процентах)

Интересным показалось посмотреть, какое выражение принимает эмоциональная компонента в зависимости от цели текста. Рассматриваемые данные можно охарактеризовать по «поджанру» в соответствии с назначенным при публикации заголовком (см. Таблицу 2): гороскопы общего характера преобладают (58,01 %), затем следуют любовные гороскопы (32,01 %) и бизнес-гороскопы (9,98 %). Как заметно по Рисунку 4, положительная лексическая тональность главным образом характерна для прогнозов в сфере любви. Обратившись к долям постов, сгруппированных по гороскопическим поджанрам и классам тональности, удалось подтвердить это наблюдение: так, 46,77 % любовных гороскопов и 42,22 % общего характера вне зависимости от знака зодиака имеют положительную «заряженность». Наиболее негативными, составляя 21,58 % постов своей подгруппы, в свою очередь являются бизнес-гороскопы (вероятно, так проявляется свойственный гороскопам этого поджанра мотив *предостережения о неприятностях на работе*), а наименее – вновь любовные (доля отрицательных постов внутри этой категории – 16,54 %).

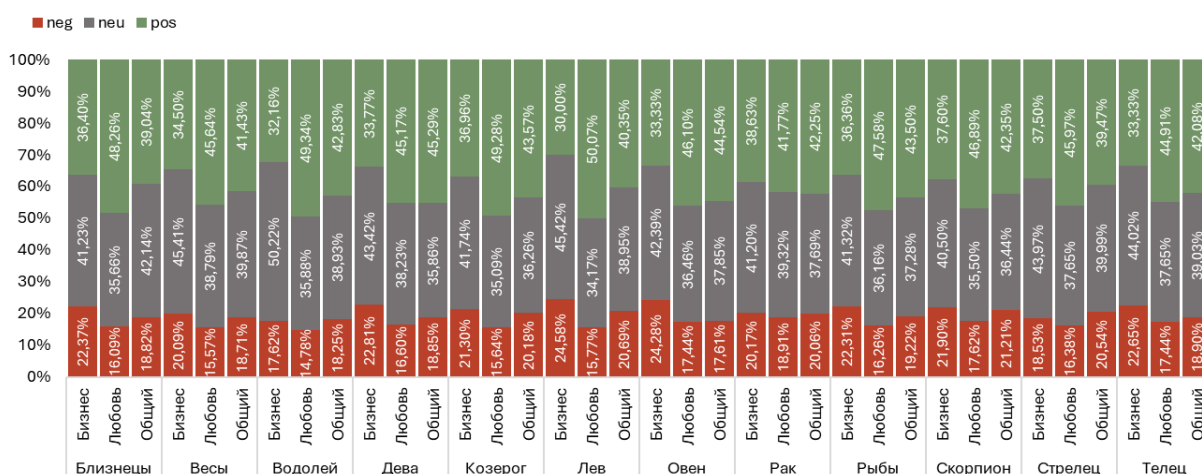


Рис. 4. Распределение гороскопов, сгруппированных в зависимости от тематической категории и подкорпуса, по тональности (в процентах)

В результате проведенного исследования были выделены лексические и дискурсивные особенности астрологических прогнозов. Частотный анализ позволяет проследить лексику, тематически соответствующую архетипическим образам знаков зодиака, однако из-за высокой генерализованности предсказаний (наблюдается значительное количество лексических повторений) и слабых различий между тематическими группами этот анализ скорее говорит о том, что гороскопы имеет смысл рассматривать как некоторый речевой жанр с маловыраженной внутренней различимостью (по крайней мере, по параметру адресата, определяемого в соответствии со знаком зодиака). Представляется, что гороскоп в современном медиаполе функционирует как «аффирмирующее послание», на что указывает также и превалирующая положительная тональность этих текстов, но не как персонализированная рекомендация.

Благодарности. Статья подготовлена в результате проведения исследования в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ).

ЛИТЕРАТУРА

1. Wuthnow R. Astrology and marginality // Journal for the Scientific Study of Religion. 1976. P. 157–168.
2. Азначеева Е. Н. Астрологический дискурс: семиотический и когнитивный аспекты // Вестник Челябинского государственного университета. 2011. № 33. С. 19–21.
3. Вепрева И. Т. Базовый лексикон астрологического прогноза: о словах с "размытым" содержанием // Экология языка и коммуникативная практика. 2017. №. 1. С. 57–63.
4. Бабаева Е. В. Лингвокультурологические характеристики русской и немецкой аксиологических картин мира: дис. ... д-ра филол. наук: Волгоград, 2004. 438 с.
5. Савицкайте Е. Р. Дискурсивные характеристики прогностических текстов (на материале немецких гороскопов): автореф. дис. ... канд. филол. Наук : Воронеж, 2006. 24 с.
6. Лалетина А. О. "Гороскоп" как гендерно-маркированный медиажанр // Жанры и типы текста в научном и медийном дискурсе. 2007. С. 288–294.
7. Мубаракшина А. М. Дискурсивные особенности гороскопического текста // Филология и культура. 2021. №. 4 (66). С. 101–105.
8. Фирак-Митц Ф. Астрологические коды. Москва : АСТ : Астрель, 2007.
9. MyStem 3.1 = морфологический парсер [Электронный ресурс] // Yandex. URL: <https://yandex.ru/dev/mystem> (дата обращения: 15.06.2025).
10. Bird S., Klein E., Loper E. Natural Language Processing with Python. O'Reilly Media Inc., 2009. URL: <https://www.nltk.org/book>.
11. Национальный корпус русского языка [Электронный ресурс]. URL: <https://ruscorpora.ru/stats> (дата обращения 01.07.2025).

12. Лукашевич Н. В., Левчик А. В. Создание лексикона оценочных слов русского языка PyСентилекс // Труды конференции OSTIS-2016. С. 377–382.
13. Orange3 Text Mining = software [Electronic resource] // Laboratory of Bioinformatics, Faculty of Computer Science, University of Ljubljana. URL: <https://orangedatamining.com> (date of access: 01.07.2025).
14. Lukashevich N. V. Automatic sentiment analysis of texts: problems and methods // Intelligent systems. Theory and applications. 2022. Vol. 26. №. 1. P. 50–61.
15. Hu, Minqing and Bing Liu (2004). Mining opinion features in customer reviews // Proceedings of AAAI Conference on Artificial Intelligence. Vol. 4. P. 755–760.
16. Колмогорова А. В., Колмогорова П. А., Куликова Е. Р. О прошлом, но в разное время: компьютерный анализ текстов учебников по истории СССР / России для шести поколений студентов // Вестник Томского государственного университета. Филология. 2024. № 89. С. 73–103. DOI:10.17223/19986645/89/4.
17. Ling Y. S. C. The use of modal auxiliary verbs in horoscope: a corpus-based study // Issues in Language Studies. 2016. Vol. 5. № 2.

УДК 81'42:32

Мальмаков Владислав Андреевич
магистрант кафедры перевода
Российский государственный педаго-
гический университет имени
А. И. Герцена
г. Санкт-Петербург, Россия

Vladislav Malmakov
Master Student of the Translation
Department
Herzen State Pedagogical University
of Russia
Saint Petersburg, Russia
vmalmakov@mail.ru

КОРПУСНЫЙ АНАЛИЗ ПОЛИТИЧЕСКОЙ РИТОРИКИ В РЕЧАХ ДОНАЛЬДА ТРАМПА И ДЖО БАЙДЕНА

В статье проводится корпусный анализ политической риторики Дональда Трампа и Джо Байдена на материале их выступлений перед Конгрессом США (2024–2025 гг.). Исследование фокусируется на выявлении лингвистических особенностей и способов речевого воздействия, таких как использование местоимений, тональность текста и частотность ключевых слов. Применяются методы морфологической и синтаксической разметки с использованием библиотек spaCy и textblob, что позволяет объективно оценить различия в риторике двух политиков. Результаты показывают, что Д. Трамп чаще использует позитивную лексику и субъективные оценки, тогда как Д. Байден делает акцент на коллективных ценностях. Исследование демонстрирует эффективность корпусных методов для анализа политического дискурса и может быть полезно для лингвистики, политологии и медиаисследований.

Ключевые слова: корпусный анализ; политическая риторика; Дональд Трамп; Джо Байден; речевое воздействие.

CORPUS-BASED ANALYSIS OF POLITICAL RHETORIC IN DONALD TRUMP'S AND JOE BIDEN'S SPEECHES

The article presents a corpus-based analysis of the political rhetoric of Donald Trump and Joe Biden, focusing on their addresses to the U.S. Congress (2024–2025). The study examines linguistic features and persuasive techniques, including pronoun usage, sentiment analysis, and keyword frequency. Methods such as morphological and syntactic tagging (using spaCy and textblob) provide an objective comparison of the two leaders' rhetorical styles. The results reveal that D. Trump employs more positive and subjective language, while J. Biden emphasizes collective values. The research highlights the effectiveness of corpus linguistics in political discourse analysis and contributes to the fields of linguistics, political science, and media studies.

Key words: corpus analysis; political rhetoric; Donald Trump; Joe Biden; persuasive techniques.

Актуальность. Исследование дополняет корпусные работы с точки зрения политической риторики.

Цель исследования. Выявить и сопоставить лингвистические особенности, способы речевого воздействия в речах Дональда Трампа (на совместном заседании Конгресса, 4 марта 2025 г.) и Джо Байдена («О положении страны», 7 марта 2024 г.) с использованием методов корпусного анализа.

Объект и предмет исследования. Объектом исследования является политическая риторика президентов США к Конгрессу как специфический жанр политической коммуникации. Предметом исследования являются лингвистические особенности Д. Трампа и Д. Байдена на материале собственного корпуса.

Политическая риторика – один из видов риторики, которая изучает речь как средство получения и осуществления государственной власти, а также оказания влияния на властные структуры [1]. Политики зачастую в своей речи прибегают к способам речевого воздействия на аудиторию. Одним из них является применение лингвистического образа «своих» и «чужих» [4, с. 36]. Они реализуются посредством инклюзивных местоимений, например в речи Д. Трампа: *We won all seven swing states, giving us an electoral college victory of 312 votes. We won the popular vote* [5]. Использование данного воздействия применимо к Д. Байдену: *We will not walk away. We will not bow down. I will not bow down* [6]. В приведенных случаях прослеживается тенденция того, что оба спикера не отделяют себя от публики, ведь они являются ее частью. Благодаря этому устанавливается положительный контакт с аудиторией, поскольку происходит непосредственное единение с народом. Что касается образа «чужих», то оба президента не сдерживаются во взаимной критике. Так, Д. Трамп заявляет следующее: *In comparison, under Joe Biden, the worst President in American history, there were hundreds of thousands of illegal crossings a month, and virtually all of them, including murderers, drug dealers, gang members, and people from mental institutions and insane asylums, were released into our country* [5]. В свою очередь, Д. Байден более сдержанным образом критикует на тот момент бывший кабинет президента Д. Трампа: *They (the last Administration) added more to the national debt than any Presidential term in American history* [6]. Данный образ строится в противостоянии республиканцев с одной стороны и демократов с другой. Американские лидеры способны оценивать ценностную систему своего народа и свободным образом опираются на нее в своих речах. Так, Д. Трамп апеллирует лозунгом *make America great again* [5], в то время как Д. Байден прибегает к применению слов *freedom, democracy, equality* [6]. Приведенные способы речевого воздействия являются лишь частью огромного множества методов влияния на аудиторию.

Корпусная лингвистика, в свою очередь, являясь разделом компьютерной лингвистики, рассматривает общие принципы построения и использования лингвистических корпусов с применением компьютерных технологий [2, с. 11]. Под языковым корпусом подразумевается большой, структурированный и размеченный массив данных, который предназначен для решения конкретных лингвистических задач. Именно поэтому корпус в области политической риторики способствует изучению речевых особенностей политических лидеров, поскольку он включает в себя несколько речей, а также позволяет определить, какими аспектами общественной жизни они апеллировали в то или иное время.

Составленный собственный корпус является одноязычным (английский язык), специализированным [3]. Он создавался на платформе Google Colab с применением встроенной библиотеки pandas (для сохранения результатов в формате Excel), встроенный модуль collections. Counter (для подсчета частотности слов), re (для фильтрации стоп-слов). Использовались сторонние библиотеки spaCy (для морфологической и синтаксической разметки) [7], textblob (для анализа тональности) [8]. Они являются сторонними, поскольку для их применения необходимо загрузить соответствующие библиотеки. Эмпирический материал включал в себя 9940 слов (речь Д. Трампа) и 8270 слов (речь Д. Байдена). Поскольку для анализа были отобраны стенограммы, слова, которые не относятся к обоим президентам, не были учтены, например, реакция публики, слова спикера Джеймса Джонсона. Применение морфологической и синтаксической разметок реализуется в рисунках 1 и 2.

```
with open(f'results/{filename}', 'w', encoding='utf-8') as f:
    for token in doc:
        f.write(f"Текст: {token.text}\n")
        f.write(f"Лемма: {token.lemma_}\n")
        f.write(f"Часть речи: {token.pos_}\n")
        f.write(f"Тег: {token.tag_}\n")
        f.write(f"Зависимость: {token.dep_}\n")
        f.write(f"Родитель: {token.head.text}\n")
        f.write("---\n")
```

Рис. 1. Код для морфологической разметки

```
with open(f'results/{filename}', 'w', encoding='utf-8') as f:
    for sent in doc.sents:
        f.write(f"Предложение: {sent.text}\n")
        for token in sent:
            f.write(f"{token.text} ({token.dep_}) -> {token.head.text}\n")
            f.write("\n" + "="*80 + "\n")
```

Рис. 2. Код для синтаксической разметки

Морфологическая разметка (part-of-speech tagging) включает не только признак части речи, но и признаки грамматических категорий, свойственных данной части речи [2, с. 36]. Разметка (annotation), в свою очередь, реализуется посредством приписывания текстам и их компонентам специальных тегов, а именно лингвистических, которые описывают лексические, грамматические и иные характеристики элементов текста, и экстралингвистических, включающих в себя информацию об авторе, название текста, год и место издания, жанр, тематику [2, с. 34]. Экстралингвистическая разметка не применялась в настоящем исследовании, поскольку она не является актуальной в отношении рассмотрения политической риторики.

В рамках корпуса код для морфологической разметки реализуются посредством команд *token.text* (определяет исходный текст токена), *token.lemma_* (определяет лемму, т. е. словарную форму слова); *token.pos_* (отвечает за часть речи, например, *VERB* – *глагол*; *token.tag_* (более детальный тег, например, *VBD* – *глагол в прошедшем времени*); *token.dep_* (определяет зависимость); *token.head.text* (определяет родительский токен). Токеном принято считать фрагмент текста, который считывается и понимается компьютером. Данный процесс проходит с каждым словом, как, например, со словом *America* в предложении *America is Back*:

Текст: America

Лемма: America

Часть речи: PROPN

Тег: NNP

Зависимость: nsubj

Родитель: is

Синтаксическая разметка состоит из команд *token.text* (текст токена); *token.dep_* (указывает тип синтаксической связи); *token.head.text* (указывает текст управляющего слова). Благодаря синтаксической разметке можно описать синтаксические связи между лексическими единицами и различные синтаксические конструкции, например, придаточное предложение, именное сказуемое и т. д. [2, с. 45]. Полученная разметка выглядит следующим образом:

Предложение Д. Трампа: *Who would want to do that?*

Who (nsubj) -> want

would (aux) -> want

want (ROOT) -> want

to (aux) -> do

do (xcomp) -> want

that (dobj) -> do

? (punct) -> want

Предложение Д. Байдена: *War was raging in Europe*

War (nsubj) -> raging

was (aux) -> raging

raging (ROOT) -> raging

in (prep) -> raging

Europe (pobj) -> in

. (punct) -> raging

Применение анализа тональности текста реализуется посредством библиотеки *textblob*. *Textblob* является библиотекой Python и применяется для обработки естественного языка (NLP), исправления орфографии, перевода.

Данная библиотека, как и spaCy, применяются в Google Colab, поскольку он основан на языке Python. Код анализа тональности текста рассмотрен в рисунке 3.

```
def analyze_sentiment(text):
    analysis = TextBlob(text)
    return {
        'polarity': analysis.sentiment.polarity, # опре
        'subjectivity': analysis.sentiment.subjectivity
    }
```

Рис. 3. Код для анализа тональности

Анализ тональности текста происходит благодаря функциям polarity (полярность) и subjectivity (субъективность). Полярность измеряется от -1 до 1, где -1 – максимально негативная лексика, 0 – нейтральная лексика и 1 – максимально положительная лексика. Приведем пример предложений с положительной и негативной оценками из риторики Д. Трампа:

Положительная оценка. [1.00] *Now we're having the best results, just about, that we've ever had. [1.00] This will be our greatest area. [0.91] I believe that my life was saved that day in Butler for a very good reason.*

Негативная оценка. [-1.00] *We suffered the worst inflation in 48 years but perhaps even in the history of our country. [-1.00] That morning, Laken was viciously attacked, assaulted, beaten, brutalized, and horrifically murdered. [-0.91] It's demeaning for women, and it's very bad for our country.*

Стоит также привести примеры таких предложений в риторике Д. Байдена:

Положительная оценка. [1.00] *We have the best economy in the world. [0.80] Here tonight is UAW President Shawn Fain, a great friend and a great labor leader. [0.70] I thank Vice President Harris for being an incredible leader defending reproductive freedom and so much more.*

Негативная оценка. [-0.75] *Hate, anger, revenge, retribution are the oldest of ideas. [-0.62] Folks at home, that's why the banks are so mad. [-0.50] Why are you angry with me providing you the opportunity for the best-educated workforce in the world?*

В отношении частотности слов применяется следующий код в рисунках 4 и 5.

```
def is_significant(token):
    if token.is_stop or token.is_punct or len(token.text) <= 2:
        return False
    if token.tag_ == "VBG" and token.pos_ == "VERB":
        return False
    unwanted_patterns = [
        r"'s$", r"n't$", r"'re$", r"'ll$",
        r"^going$", r"^gonna$", r"^wanna$",
        r"^would$", r"^could$", r"^should$"
    ]
    allowed_pos = {"NOUN", "PROPN", "ADJ", "VERB"}
    return (
        token.pos_ in allowed_pos and
        not any(re.search(patt, token.text.lower()) for patt in unwanted_patterns)
    )
```

Рис. 4. Код для подсчета частотности слов

```
def get_top_words(doc, n=15):
    words = [token.lemma_.lower() for token in doc if is_significant(token)]
    return Counter(words).most_common(n)

top_trump_words = get_top_words(doc_trump)
top_biden_words = get_top_words(doc_biden)
```

Рис. 5. Продолжение кода для подсчета частотности слов

Прежде чем подсчитать частотность слов посредством модуля *re*, необходимо задать ряд условий (рисунок 4). Первое условие не принимает в учет слова, которые являются стоп-словами, знаком пунктуации или если длина токена равна или меньше двух. Второе условие не учитывает глаголы в форме герундия или причастия. Третьим условием выступают сокращения (*'s*, *n't*, *'re*, *'ll*), разговорные формы (*going*, *gonna*, *wanna*), модальные глаголы (*would*, *could*, *should*). В случае если не реализуется какое-либо из условий, происходит подсчет частотных значимых слов (15 слов) в нижнем регистре (*token.lemma.lower*).

Результаты исследования

Получены результаты размеченных речей обоих президентов с точки зрения морфологии и синтаксиса. В отношении тональности, речь Д. Трампа стремится к субъективизму (0,49) и употреблению более положительной лексики (0,16) чем Д. Байден (0,42) и (0,14) соответственно. Самыми частотными словами у Д. Трампа оказались *thank* (69), *country* (48), *people* (47) и т. д. У Д. Байдена – *America* (44), *year* (40), *american* (34) и т. д. Такая частотность слов отображает стремление обоих политиков заручиться большей поддержкой со стороны американского народа, а также их стремление не

отдаляться от него. Данные результаты применимы для анализа особенностей риторики обоих американских президентов, а также для статистических данных.

Заключение

Корпус применим для проведения исследования по таким лингвистическим вопросам, как морфология и синтаксису, для анализа тональности текста и определения частотности значимых слов в риторике Д. Трампа и Д. Байдена во время обращения к Конгрессу США. Для наиболее подробного исследования настоящего вопроса возможно дальнейшее рассмотрение иных речей президентов, а также рассмотреть особенности применения таких риторических приемов, как метафоры, сравнения (*simile*), эпитеты.

ЛИТЕРАТУРА

1. Данилина В. В. Политическая риторика как научное направление // Вестник Московского университета. Серия 21. Управление (государство и общество). 2005. №2. URL: <https://cyberleninka.ru/article/n/politicheskaya-ritorika-kak-nauchnoe-napravlenie> (дата обращения: 10.06.2025).
2. Захаров, В. П. Корпусная лингвистика : учебник / В. П. Захаров, С. Ю. Богданова. – 3-е изд., перераб. – Санкт-Петербург : СПбГУ, 2020. – 234 с. – ISBN 978-5-288-05997-1. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/139576> (дата обращения: 10.06.2025).
3. Собственный корпус [Электронный источник]. URL: <https://colab.research.google.com/drive/1yFj56EkEh8PAMMbCmpqufexPRAEKJVbR?usp=sharing> (дата обращения: 11.06.2025).
4. Чернявская В. Е, Молодыхенко Е.Н. История в дискурсе политики: лингвистический образ «своих» и «чужих». Изд. Стереотип. – М.: ЛЕНАНД, 2018. – 200 с.
5. Donald J. Trump. 5th Article II Sec. 3 Address to a Joint Session of Congress [Электронный ресурс]. Режим доступа: <https://www.americanrhetoric.com/speeches/stateoftheunion2025.htm> (дата обращения: 11.06.2025).
6. Joe Biden. Third Presidential State of the Union Address [Электронный ресурс]. URL: <https://www.americanrhetoric.com/speeches/stateoftheunion2024.htm> (дата обращения: 11.06.2025).
7. SpaCy. Industrial-Strength Natural Language Processing [Электронный ресурс]. URL: <https://spacy.io/> (дата обращения: 11.06.2025).
8. TextBlob: Simplified Text Processing [Электронный ресурс]. URL: <https://textblob.readthedocs.io/en/dev/> (дата обращения: 11.06.2025).

УДК 81 '32:81 '365:81 '374:81 '37

Мацегора Маргарита Валентиновна

ассистент

Приазовский государственный

технический университет

г. Пятигорск, Россия

Margarita Matsegora

Teaching assistant

Priazovsky State Technical

University

Pyatigorsk, Russia

professor9016@gmail.com

**ГИПЕРБОЛА И ПАРАДОКС КАК КОГНИТИВНЫЕ МЕХАНИЗМЫ
ПЕРЕОСМЫСЛЕНИЯ КОНЦЕПТА HUMANITY
В ДИСКУРСЕ ПОЛИТИЧЕСКИХ И ГУМАНИТАРНЫХ ОРГАНИЗАЦИЙ
(на материале современного английского языка)**

Статья посвящена изучению вербальной репрезентации концепта HUMANITY в условиях социальной нестабильности и идеологического противостояния. Анализ эмпирического материала показывает, что языковые стратегии, основанные на гиперболизации и парадоксальности высказывания, не только придают ему стилистическую окрашенность, но и выступают когнитивными механизмами, способствующими переосмыслению ценностных категорий в политическом и гуманитарном дискурсах. Основной целью статьи является выявление функций гиперболы и парадокса как механизмов воздействия на общественное сознание с высоким суггестивным потенциалом. Установлено, что данные явления активно используются в политическом и гуманитарном дискурсах для усиления эмоционального восприятия и формирования альтернативных интерпретаций реальности. Теоретической основой послужили исследования, подчеркивающие, что парадоксальность высказываний в политической риторике часто отражает столкновение ценностных систем или используется для обесценивания устоявшихся моральных норм. Гиперболизация, в свою очередь, позволяет формировать у аудитории эмоциональный отклик, важный для реализации политической повестки. На примерах из речей политиков и представителей гуманитарных организаций показано, что гиперболизация усиливает значимость информации и мотивирует слушающего, а парадокс запускает рефлексию и приводит к пересмотру общепринятых норм. В заключении делается вывод о том, что применение указанных стратегий позволяет продуцентам дискурса переосмысливать концепт HUMANITY, трансформировать устоявшиеся смыслы и конструировать новые модели вербализации гуманности как одной из ключевых онтологических категорий современного общества.

Ключевые слова: концепт HUMANITY; гипербола; парадокс; политический дискурс; гуманитарный дискурс; ценностные ориентиры; языковое воздействие.

**HYPERBOLE AND PARADOX AS COGNITIVE MECHANISMS
FOR RETHINKING THE CONCEPT OF HUMANITY
IN THE DISCOURSE OF POLITICAL AND HUMANITARIAN ENTITIES
(Based on Modern English Language Material)**

The article is devoted to the study of the verbal representation of the concept HUMANITY under conditions of social instability and ideological confrontation. Analysis of empirical data shows that linguistic strategies based on hyperbolization and paradoxicality of speech not only impart stylistic coloring to the concept, but also function as cognitive mechanisms that contribute to the re-evaluation of value categories in political and humanitarian discourse. The main aim of the article is to identify the functions of hyperbole and paradox as mechanisms of influence on

public consciousness with high suggestive potential. It has been established that these phenomena are actively used in political and humanitarian discourse to intensify emotional perception and construct alternative interpretations of reality. The theoretical basis includes studies that emphasize how the deliberate paradoxicality of statements in political rhetoric often reflects a clash of value systems or is used to devalue previously significant moral norms. Hyperbolization, in turn, helps generate an emotional response in the audience, crucial for advancing a political agenda. Using examples from politicians' speeches and statements by representatives of humanitarian organizations, it is shown that hyperbole intensifies the perceived importance of information and motivates the listener, while paradox initiates reflection and leads to a reconsideration of established norms. The conclusion emphasizes that the use of these strategies allows discourse producers to reinterpret the concept of HUMANITY, transform established meanings, and construct new models for verbalizing humanity as one of the key ontological categories in modern society.

Key words: concept of HUMANITY; hyperbole; paradox; political discourse; humanitarian discourse; axiological landmarks; persuasion.

В условиях социальной неопределенности и эскалации конфликтов язык приобретает все большее значение как инструмент формирования ценностных ориентиров общества. Особую значимость в этом процессе приобретают дискурсы политических институтов и гуманитарных организаций, где универсальные принципы, связанные с человечностью, все чаще становятся предметом дискуссий и идеологических споров. Вербальная репрезентация концепта HUMANITY в таких условиях характеризуется повышенной экспрессивностью и оценочностью, а языковые стратегии, направленные на усиление или парадоксальное переосмысление значений, выступают эффективным средством воздействия на общественное сознание [5, с. 126].

Такие стратегии, как гиперболизация и парадоксализация, являются не просто стилистическими приемами, но и важными механизмами, способствующими переосмыслению ключевых понятий в публичной коммуникации. Гипербола и парадокс представляют собой ярко выраженные стилистические явления, выполняющие важную роль в организации речевой коммуникации. Гипербола, как экспрессивный прием, предполагает значительное [6, с. 68] преувеличение качественных или количественных характеристик объекта, что способствует акцентуации авторской позиции и усиливает эмоциональное воздействие на адресата [1, с. 196]. Этот механизм активно используется в устной и письменной речи для создания яркого образа, формирования оценочного отношения или выделения ключевых смысловых доминант. Парадокс, в свою очередь, заключается в формулировании высказывания, противоречащего общепринятым представлениям или логическим ожиданиям, что вызывает у слушателя эффект неожиданности и побуждает к рефлексии. [4, с. 543] Такой прием позволяет поставить под сомнение устойчивые нормы восприятия и предлагает новую интерпретацию известных фактов или понятий. Оба явления – гипербола и парадокс – выходят за рамки чисто стилистической функции, становясь эффективными инструментами дискурсивного влияния, способными модифицировать восприятие действительности и формировать новые смысловые установки [2, с. 106].

Так, данные стилистически маркированные коммуникативно-лингвистические явления выступают не просто элементами стилистической выразительности, но и важными когнитивно-дискурсивными механизмами, способствующими переосмыслению ценностных категорий в условиях социальной неопределенности. В современной лингвистике эти явления рассматриваются как эффективные средства воздействия на общественное сознание, особенно в политическом и гуманитарном дискурсах, где они выполняют функцию актуализации ключевых смысловых доминант. Н. Л. Харченко подчеркивает, что «парадоксализация способна нарушать ожидания адресата, вызывая рефлекссию и формируя новые интерпретационные модели в условиях обострения идеологических противоречий» [7, с. 177]. Аналогичную позицию занимает и А. А. Марьянова, которая утверждает, что в спонтанной речи политиков парадоксальные формулировки нередко возникают как результат импровизации под давлением момента, что может приводить как к созданию новых, неожиданных смысловых акцентов, так и к коммуникативным неудачам [4, с. 543]. Автор отмечает, что такие высказывания способны не только вызвать интерес и обсуждение, но и изменить восприятие политического события, формируя альтернативную интерпретацию реальности. При этом парадоксализация в данном случае часто выступает не как осознанная риторическая стратегия, а как побочный эффект эмоциональной и когнитивной нагрузки на речь политика в условиях публичного выступления. Как показывает анализ, представленный в работе *Policy Paradox and Policy Prospects: Editors' Introduction* / 'Парадоксы политики и ее перспективы: Вступительная статья редакторов', парадоксы в политике часто возникают не случайно, а являются результатом столкновения конкурирующих ценностей, норм и интересов, что делает их органичной частью публичного дискурса. Авторы подчеркивают, что такие формулировки служат важным инструментом формирования общественного мнения, поскольку позволяют говорящему представить сложные ситуации в новом свете – через призму противоречия [10].

Этот подход перекликается с фундаментальной работой Деборы Стоун *Policy Paradox* / 'Политический парадокс', где она демонстрирует, что политические споры зачастую строятся не вокруг фактов или интересов, а вокруг конфликтующих ценностей, которые могут быть осознанно выражены через парадоксальные формулировки. Например, когда в речи встречается выражение вроде *humanity has become a luxury in times of survival* / 'человечность стала роскошью во времена выживания', это не просто языковая игра, а отражение реального напряжения между гуманистическими принципами и практической необходимостью [9]. Такие высказывания не только привлекают внимание, но и побуждают к переосмыслению сложившихся установок, особенно в условиях кризисов, связанных с человечностью и правами человека.

Гиперболизация также получила широкое освещение в научной литературе. Н. Н. Левнина анализируя речевые стратегии государственных деятелей, демонстрирует, что преувеличение служит мощным инструментом усиления оценочности и формирования эмоционального фона, особенно в кризисных ситуациях [3, с. 132]. За рубежом данную тенденцию подтверждает М. А. Сали, который в рамках анализа английских политических речей демонстрирует, что гиперболизация выступает эффективным риторическим средством, направленным на усиление воздействия на аудиторию. Автор отмечает, что политики сознательно используют преувеличение, чтобы представить ситуации, явления или лица более значимыми, чем они есть на самом деле, тем самым формируя эмоциональный отклик и мобилизуя слушателей вокруг определенной идеи или цели [8, с. 47]. Таким образом, как показывают исследования, гипербола и парадокс становятся не просто экспрессивными средствами, но важными механизмами вербального влияния, которые активно задействуются в коммуникации, ориентированной на переосмысление универсальных ценностей, включая концепт HUMANITY.

Примеры из речей политиков и брифингов подтверждают, что гиперболизация и парадоксализация активно используются как стратегии усиления эмоционального воздействия и формирования оценочных установок. Эти приемы позволяют говорящему не просто передать информацию, но и модифицировать восприятие реальности, делая ее более символически насыщенной и эмоционально окрашенной. Анализ текстов показывает, что данные языковые явления часто становятся ключевыми элементами вербализации концепта HUMANITY, особенно в условиях кризиса, социального напряжения или идеологического противостояния.

Гиперболизация, как уже отмечалось, заключается в сознательном преувеличении качественных или количественных характеристик объекта с целью акцентуации авторской позиции.

Так, в высказывании ниже (1) слово *humongous* выступает как одиночная гипербола, усиливающая масштаб проблемы и создающая эффект эмоциональной вовлеченности. Такой прием позволяет выделить проблему как исключительную и требующую немедленного внимания:

(1) *The mental health crisis is humongous* / Кризис психического здоровья – колоссальный [13].

Многократное употребление глагола *steal* в сочетании с нарастанием значимости объектов (*jobs – wealth – country*) образует градуированную гиперболу, которая не только преувеличивает угрозу, но и усиливает негативную оценку в адрес группы, обозначенной как «они». Подобная структура часто используется для мобилизации аудитории вокруг определенной идеи или политической повестки:

(2) *They steal our jobs, they steal our wealth, they steal our country* / ‘Они отнимают у нас рабочие места, лишают нас благополучия, похищают нашу страну.’

В следующем примере гиперболизация строится на принципе контекстуального преувеличения, где отсутствие финансовых ресурсов у организации представляется как полная невозможность ее существования. Это способствует созданию образа внешнего врага, который существует вопреки всему, и, следовательно, является особенно опасным:

(3) *You wouldn't have had – Hamas had no money. Do you know that? / 'Разве вы не понимаете – у ХАМАСа не было средств!' [13].*

Парадоксализация, в свою очередь, выражается посредством семантически и прагматически противоречивых высказываний, нарушающих существующие пресуппозиции и социальные конвенции, следствием чего является активизация ментальных процессов логического следования и инференции, призванных «высветить», экстериоризировать имплицитные смысловые элементы для преодоления смысловой нелогичности или прагматической адекватности высказывания. Как правило, данные элементы извлекаются из лингвистического и ситуативного (экстралингвистического) контекстов.

Так, трехкратное повторение одного и того же утверждения, оформленное как стилистическая анафора, создает эффект парадокса – чем больше говорится о том, чего не хотят, тем более вероятным становится это действие в воображении адресата. Такое построение высказывания может быть интерпретировано как невербальная компенсация за возможные действия, что указывает на двойственность позиции говорящего:

(4) *We do not seek – we do not seek – we do not seek to have American troops fighting in Russia or against Russia / 'Мы не стремимся – мы не стремимся – мы не стремимся к тому, чтобы американские войска сражались в России или против России [12].'*

Повтор слова *still* после упоминания сложных исторических обстоятельств создает парадокс между устойчивостью идеала и реалиями современности, заставляя задуматься о напряженности между образом и действительностью. Такой прием характерен для риторики лидеров, стремящихся сохранить легитимность своей власти в условиях кризиса:

(5) *America is a beacon to the world, still, still / 'Америка остается маяком для мира, остается, остается [12].'*

Абсолютизация принадлежности («вы все – Америка») противоречит социальной реальности, где существует разделение на «своих» и «чужих». Это высказывание работает как парадоксальная декларация единства, вызывающая переосмысление понятий идентичности и принадлежности. Такие формулировки могут быть результатом импровизации в условиях публичного давления, но они приобретают самостоятельный смысловой вес, если находят отклик в общественном сознании:

(6) *You're all America. You're all America / 'Вы все – Америка. Вы все – Америка [12].'*

Примеры из выступлений представителей гуманитарных организаций также демонстрируют активное использование гиперболизации и парадокса-

лизации как средств усиления этической оценки и формирования общественного внимания к вопросам человечности. В отличие от политических речей, где доминируют идеологические установки, в данном случае акцент делается на страдании, необходимости помощи и моральной ответственности.

Один из показательных случаев использования гиперболизации можно найти в высказывании Джанти Соэрипто (Janti Soeripto), президента организации *Save the Children*:

(7) *It is horrific. And it is – and people stop hearing, I think, when they hear all those numbers. Thousands. And they forget that behind every – all those thousands, there is a child and their families behind it* / ‘Это ужасающе. И люди перестают слышать, я думаю, когда слышат все эти цифры. Тысячи. И они забывают, что за каждой – за всеми этими тысячами стоит ребенок и его семья [11].

Здесь слово *horrific*, повторяющееся дважды, служит не только экспрессивным маркером, но и усилителем эмоционального восприятия масштаба трагедии. Фраза *they forget that behind every – all those thousands, there is a child...* содержит элемент градуированной гиперболы, направленной на переосмысление статистики через личный опыт. Это позволяет говорящему перенести акцент с безликих цифр на индивидуальные судьбы, тем самым актуализируя ценность человеческой жизни.

Еще один случай парадоксализации содержится в выступлении Филиппо Гранди, Верховного комиссара ООН по делам беженцев:

(8) *The next campaign for inclusion will be tougher, and it is inclusion in social and economic responses ...* / ‘Следующая кампания за инклюзию будет сложнее, и это инклюзия в социальные и экономические программы...[14].’

Формулировка *the next campaign for inclusion will be tougher* звучит парадоксально, потому что включение, или инклюзия, обычно ассоциируется с прогрессом, развитием и улучшением положения дел. Однако здесь подчеркивается, что инклюзия требует все большего усилия, что противоречит общепринятому позитивному восприятию этого понятия. Такой прием заставляет адресата задуматься о реальных трудностях реализации гуманитарных принципов в условиях экономического и политического кризиса.

Эти примеры еще раз подтверждают, что гипербола и парадокс выполняют не только экспрессивную функцию, но и играют важную роль в формировании общественного мнения и интерпретации универсальных ценностей, таких как HUMANITY. Эти механизмы позволяют говорящему не просто передать сообщение, но и направить его восприятие в нужное русло, тем самым модифицируя ценности и установки аудитории. Как показывает практика политической коммуникации, особенно в условиях кризиса, язык приобретает не только информативную, но и перформативную силу, превращаясь в инструмент формирования новой реальности.

ЛИТЕРАТУРА

1. Голубева Т. М. Персуазивность гиперболы в политическом дискурсе (на материале высказываний британских политиков) // Вестник Волгоградского государственного университета. 2019. № 3. С. 195–205.
2. Иванова Т. К. Средства экспрессии в политическом медиадискурсе и опасность их использования // Российские исследования. 2021. № 2. С. 106–112.
3. Левина Н. Н. Гипербола в политическом дискурсе // Университетские чтения – 2017: Материалы научно-методических чтений, Пятигорск, 12–13 января 2017 года. Ч. III. Пятигорск : Пятигорский государственный лингвистический университет, 2017. С. 132–136. EDN ZFZZDB.
4. Речевые оплошности и коммуникативные неудачи в спонтанной речи политиков / А. А. Марьянова, А. М. Чапанова, И. С. Назарова, Е. Д. Шамсудинова // Экономика и социум. 2022. № 6–2 (97). С. 541–546.
5. Тищенко С. В. Феномен речевого воздействия в современном политическом дискурсе: тактики, стратегии и языковые средства их актуализации / С. В. Тищенко, Н. М. Джандубаева // Университетские чтения – 2021 : Материалы научно-методических чтений ПГУ, Пятигорск, 17–19 марта 2021 года. Том Часть II. – Пятигорск: Пятигорский государственный университет, 2021. – С. 126–131. – EDN FTXDQS.
6. Трифонова С. А. Гиперболизация как стилистический приём создания эвфемизмов // Вопросы современной лингвистики. 2018. № 4. С. 68–74.
7. Харченко Н. Л. Языковые средства выражения элементов фантастики в политическом дискурсе // Гуманитарные науки. 2021. С. 176–179.
8. Salih S. M., Braim M. M. A Textual Analysis of Hyperbole and Litotes in Selected English Political Speeches [Electronic Resource] // Anbar University Journal of Languages & Literature. 2014. № 16. P. 45–67. URL: file:///C:/Users/User/Downloads/Policy_Paradox_and_Policy_Proprospects_Editors_Intro.pdf (date of access: 28.06.2024).
9. Stone D. Policy Paradox: The Art of Political Decision-Making [Electronic Resource] 3rd ed. Wiley, 2012 [1988]. URL: file:///C:/Users/User/Downloads/Deborah_Stone_Policy_Paradox_The_Art_of_Political_.pdf (date of access: 28.06.2024).
10. Zahariadis N., Zittoun P. Policy Paradox and Policy Prospects: Editors' Introduction to the Keynote Speech by Deborah Stone [Electronic resource] // International Review of Public Policy. 2024. Vol. 6. № 1. URL: <https://journals.openedition.org/irpp/4082> (date of access: 28.06.2025).
11. Transcript: Avril Benoit, Doctors Without Borders & Janti Soeripto, Save the Children on "Face the Nation" [Electronic resource] // CBS News. 2024, April 7. URL: <https://www.cbsnews.com/news/avril-benoit-doctors-without-borders-janti-soeripto-save-the-children-face-the-nation-transcript-04-07-2024/> (date of access: 28.06.2025).
12. Transcript of Biden's Speech on Israel and Ukraine [Electronic resource] // The New York Times. 2023, October 19. URL: <https://www.nytimes.com/2023/10/19/us/politics/transcript-biden-speech-israel-ukraine.html> (date of access: 28.06.2025).

13. Discussion on the War in Gaza at World Economic Forum [Electronic resource] // Rev.com. 2024. URL: <https://www.rev.com/blog/transcripts/discussion-on-the-war-in-gaza-at-world-economic-forum> (date of access: 28.06.2025).
14. World Health Organization (WHO) Coronavirus Press Conference [Electronic resource] // Rev.com. 2020, June 19. URL: <https://www.rev.com/transcripts/world-health-organization-who-coronavirus-press-conference-june-19> (date of access: 28.06.2025).

УДК 80.81

Мустафаева Сахила Багхир
кандидат филологических наук,
доцент
Азербайджанский университет
языков
г. Баку, Азербайджан

Sahila Baghir Mustafayeva
Ph.D. in Philology,
Associate Professor
Azerbaijan University of Languages
Baku, Azerbaijan
mustafayeva.sahila@gmail.com

ДИСКУРС И КОРПУСНЫЕ ИССЛЕДОВАНИЯ В СОВРЕМЕННОЙ ЛИНГВИСТИКЕ

Статья посвящена дискурсу и корпусной лингвистике в современной лингвистике. Проанализировано значение термина «дискурс». Под дискурсом подразумевается совокупность более двух предложений, связанных между собой смыслом и содержанием. Семантические отношения между предложениями, составляющими дискурс, рассматриваются как единая структура, объединяемая в рамках изучения дискурса. Проанализированы мнения известных лингвистов, изучающих дискурс и корпусную лингвистику. В статье подчеркивается важность различий между дискурсом и текстом для лингвистики. Таким образом, текст и дискурс рассматриваются в контексте взаимосвязи, каждый из которых раскрывает свои особенности. Отмечается, что социальные последствия играют ключевую роль в изучении дискурса. Текст представляет собой абстрактную ментальную конструкцию, наделенную свойствами семантической целостности и грамматической полноты, хотя дискурс является атрибутом коммуникативно-социальной сферы и текстового сознания. В статье также рассматриваются такие термины, как референция, пропозиция, импликация, вывод и т. д. Подход Ф. Вейсалли в этой области подробно рассматривается. В статье обращается внимание на взаимосвязь между знаком и материальным миром – референтные отношения; референтные отношения рассматриваются как обязательные для учета в дискурс-анализе. В статье корпусные исследования рассматриваются через призму дискурса. Рассматриваются корпуса и их использование для получения объективных, количественных и эмпирических данных. Корпусные данные рассматриваются как альтернативный инструмент анализа данных в формальной лингвистике. Корпусная лингвистика как методологический инструмент рассматривается как признанное и широко используемое в исследованиях направление в лингвистике, и ее ценность в этой области высоко оценивается. В статье также излагается собственная точка зрения автора на изучение дискурса в корпусной лингвистике.

Ключевые слова: дискурс; текст; взаимосвязь; связность; корпус; данные.

DISCOURSE AND CORPUS STUDIES IN MODERN LINGUISTICS

The article deals with discourse and corpus studies in modern linguistics. The meaning of the term discourse has been analyzed. Discourse is meant to include more than two sentences and these sentences are related to each other through thought and content. The semantic relationships between the sentences that are used to make up a discourse are presented to be combined into a single structure in the study of discourse. The opinions of famous linguists have been analyzed in the study of discourse and corpus linguistics. The differences between discourse and text are highlighted to be very important in linguistics in the article. So, text and discourse are explained in terms with each other; each covering its own features. Social consequences are mentioned to be crucial in the study of discourse. The text is a kind of abstract mental construction endowed with the properties of semantic integrity and grammatical completeness though discourse is an attribute of the communicative-social sphere, and text

consciousness. The terms such as *reference*, *proposition*, *implication*, *inference*, etc. are also touched upon in the article. The approach of F. Veysalli has been written about in this field. The article draws attention to the relationship between the sign and the material world – the referent relationship; referent relationship is written about to be taken into account in discourse analysis. The corpus studies have been explained through discourse in the article. It writes about corpora and its usage in order to provide objective, quantitative and empirical evidence. Corpus data is meant to be an alternative tool for analyzing data in formal linguistics. Corpus linguistics as a methodological tool is explained to be accepted as a well-established field in linguistics and is widely used in research and its value in this field is highly appreciated. The author's own point of view in the study of discourse in corpus linguistics has been written about in the article as well.

Key words: discourse; text; relationship; coherence; corpus; data.

The term discourse in linguistics refers to a combination of two or more sentences that are related to each other in terms of thought and content. Due to its place in the syntax system, discourse is a language-speech unit that is larger than a sentence but smaller than a text. There is both a semantic relationship between the sentences that make up a discourse, and they are combined into a single structure. As a syntactic whole, discourse is located between the sentence and the text, connecting them.

Human consciousness, thinking, and worldview are formed under the influence of the discourses that surround them in everyday life in the modern world. Discourse applies to all areas of an individual's activity. It can be represented by any written or oral expressions, for example: scientific publications and educational materials, lectures, speeches of journalists, politicians, and cultural figures in the periodical press and media, in business correspondence, in the classroom, etc.

There are many linguists who are known to have studied discourse as well as text in linguistics.

A. Mammadov and M. Mammadov characterize the main differences between discourse and text as following:

- 1) Discourse is pragmatic; text is an exclusively linguistic category;
- 2) Discourse is a process category; text is a result category (static);
- 3) Text is an abstract construct; discourse is its actualization [1, p. 41].

In more detail, these differences are defined to be the following:

It occurs where there are social consequences of the discussion. The text is a kind of abstract mental construction endowed with the properties of semantic integrity and grammatical completeness though discourse is an attribute of the communicative-social sphere, and text consciousness.

It can be determined that the written and oral language users encounter various differences in the use of language. The speaker can increase the impact of the words that are used with the help of various paralinguistic means, for example, facial expressions, gestures, etc.; such paralinguistic means are not used in writing though.

The dependence of a word and a sentence on the word or sentence that comes before and after it as well as the study of their relationships with other objects and real reality, played an important role in traditional linguistics. The metalinguistics of these aspects necessitates the development of a number of specific terms.

Discussion

As we have noted, the pragmatic approach to the study of language in terms of use is known to be more important. In other words, the formal approach is not so important in this approach. In order to clarify information in the area where discourse occurs, deictic forms, for example, /here, there, I, you, this, that, etc. / are used, but in addition to deixis, the terms *reference*, *proposition*, *implication*, *inference*, etc. can also be attributed to this. According to J. Lyons, the relationship between a word and an object can be called reference. This traditional view shows itself as a linguistic expression of the relationship between language and the world. Later, J. Lyons also puts forward the following idea: the speaker refers (with various appropriate expressions); he/she discovers the necessary expression through the referent [2, p. 67]. In this regard, P.F. Strawson [3, p. 54] argues that reference is not something that the expression does; this is a means by which the speaker uses the appropriate expression. Regarding this idea, J. Searle's view is as following: expressions do not express anything other than giving a promise or giving an order [4, p. 98]. According to F. Veysalli, the relationship between the sign and the material world - the referent relationship – must be taken into account in discourse analysis. Thus, in discourse analysis, the referent is understood as the one who acts on the part of the speaker or writer [5, p. 54].

The study of coherence in discourse is one of the most important factors. As A. Mammadov and M. Mammadov also noted, apart from discourse coherence, no mutual meaning relationship between components has the opportunity to be developed in a clear, isolated form from the others. Other types of meaning are always added to one meaning relationship to one degree or another [1, p.25]. They also note that in multi-component discourse texts, it is not necessary for any meaning relationship to manifest itself from the beginning to the end of the text as different types of meaning relationships can occur within the discourse.

All these issues of coherence help to define the proposition. The proposition is also defined to carry main features in this respect as following: "*A proposition is an idea that the speaker has obtained or is able to grasp within the general topic during the conversation of the participants*". It should be noted that in both of the quotes we have shown above, the main source of the proposition is a speaker.

Corpus linguistics

Applied linguistics is known to be benefited greatly from corpus studies. Most formal grammatical rules were previously described as dependent on intuition [6, p. 16]. This way of studying grammar became especially popular after N. Chomsky's claim that grammatical rules are *abstract rational* [7, p. 90]. However, with the rise of corpus linguistics after the 1980s, researchers such as J. Sinclair suggested that empirical data were needed to prove abstract rules [8, p. 20]. As a result, corpora began to be used to provide objective, quantitative and empirical evidence, and corpus data was an alternative tool for analyzing data in formal linguistics. Today, corpus linguistics as a methodological tool is accepted as a well-established field in linguistics and is widely used in research and its value in this field is highly appreciated.

Some advantages and some disadvantages of corpus linguistics have been identified.

First, corpus linguistics provides a wealth of information for language research by providing many examples of a particular structure, and linguists can work out how the structure is used by looking at this information. Corpus linguistics is a really good tool for collecting data for language research and can provide a large amount of information that would be very time-consuming with traditional data collection methods.

Second, in corpus-based linguistic research, interpretations and evaluations of language use are not accepted as reliable evidence because they are completely subjective and difficult to empirically verify [9, p. 93]. Therefore, empirical data form the basis of corpus linguistics and, unlike qualitative methods, encourage researchers to avoid subjective interpretations or opinions during analysis. In this case, the data obtained from the corpus are analyzed statistically and concrete evidence is used. Therefore, corpus linguistics depends on objective and empirical data, and the data obtained from it are subject to scientific and objective verification. The third advantage of corpus linguistics is that it provides some unified information about language structures.

S. Mustafayeva highlights the usage of intonation patterns, their usage of frequency in speech and the role of context in corpus linguistics [10, p. 739].

The statistical software used in corpus linguistics provides frequencies and statistics about certain structures in a study. For example, when a language study is conducted on the use of quantitative indicators, corpus linguistics can provide frequencies of use, through which researchers can get acquainted with some rules about these structures. In other words, corpus linguistics does not focus on single contexts, but rather provides a general picture of a phenomenon, thus allowing researchers to have a general understanding of the nature of that particular phenomenon. Thus, these results are considered relevant and reliable in language research, because they are taken from tens of thousands of uses of real-life data. In this regard, corpus linguistics should be considered truly successful in large-scale analysis.

Another advantage of corpus linguistics is that it contains basic knowledge about the information it offers. Most modern corpora include information about the socio-economic status, education, gender, or birthplace of speakers. In addition, corpora can also provide information by topic, field, or genre. For example, a researcher studying bilinguals can learn about bilingual speakers by looking at their background, or a sociolinguist studying sociolinguistic factors in language can learn about the socioeconomic status of speakers in a corpus. This is another important advantage of corpus linguistics, as researchers can gain relevant knowledge from the information included in the corpus about speakers or topics.

As for the shortcomings of corpus linguistics data analysis, first of all, corpus linguistics is strictly quantitative and relies on the description of language through numbers and statistics. Although single statistics can be valuable in language research, corpus linguistics does not capture some of the most important

components of language, such as the context and structure of speech between speakers. Language is a social phenomenon and it is difficult to separate it from its context [9, p. 15]. Corpus linguistics research can tell researchers how often a structure is used, but it cannot tell when, where, or how it is used. This is a real problem for those who want to use corpus linguistics in qualitative language research.

Furthermore, corpus linguistics does not consider intuition as a means of gathering information, as it is highly subjective and untestable. This could be seen as a shortcoming of corpus linguistics, but it is important in terms of competence and performance in language research. It is also considered a reliable method in terms of studying examples taken from nature, representing the language performance of speakers, but researchers such as M. Nieuwland and A. Martin [10, p. 102] suggest that performance does not equal competence. In their opinion, what is in the minds of speakers is also relevant and indeed necessary for language learning.

Discourse is different from text; it has its own characteristics. Discourse is a broader concept than text.

Corpus linguistics involves a large database.

It (corpus linguistics) focuses on large-scale analysis and is good at providing descriptive statistics; it is weak at offering detailed analysis of individual users. What is most important for corpus linguistics is not what individual speakers do, but what the “whole” data suggests. This is of course a problem for corpus linguistics as it cannot account for individual variation in language use. It can provide information based on the background of speakers, and while that information may be generalizable about certain groups, it is still at the group level and does not include individual contexts.

REFERENCES

1. Mammadov, A., Mammadov, M. Cognitive perspectives of discourse analysis. Baku: Chasioğlu. 2010, 96 p.
2. Lyons J. Introduction to Theoretical Linguistics. Cambridge: Cambridge University Press. 1968, 290 p.
3. Strawson P.F. On referring Mind. Oxford: Oxford Universety Press, 1959. p. 45–55.
4. Searle J.R. Speech Acts. Cambridge: Cambridge University Press, 1969.203 p.
5. Veysalli F.Y. Introduction to Discourse Analysis. BakU: “Education” NPM, 2010. 114 p.
6. Aarts, J. Intuition-Based and Observation-Based Grammars. In Aijmer, K and B. Chomsky, N. Syntactic structures. Mouton: The Hague, 1957.135 p.
7. Sinclair, J. M. Corpus, Concordance and Collocation. Oxford: Oxford University Press, 1991.197 p.
8. Kennedy, G. An Introduction to Corpus Linguistics. London: Pearson Education Limited. 328 p.

9. Mustafayeva, S.B. Corpus linguistics as a new field // University of Science and Technologies and International scientific journal “Science and Innovation” for participation in the International scientific and practical conference “Modern philological issues: problems and solutions”. 2025, Number 1. p. 739–742.
10. Duranti, A., and Goodwin, C. (Eds.). Rethinking context. Language as an interactive phenomenon. Cambridge: Cambridge University Press, 1992. 209 p.
11. Nieuwland, M. S. and Martin, A. E. (2012). If the real-world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, 122, 102–109.

УДК 811.131.1

Овсейчик Юлия Владимировна
кандидат филологических наук,
заведующий кафедрой теорети-
ческой и прикладной лингвистики
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Yuliya Auseichyk
PhD in Philology, Head of the
Theoretical and Applied Linguistics
Department
Belarusian State University of Foreign
Languages
Minsk, Belarus
ovsei77@rambler.ru

СТАТИСТИЧЕСКИЕ ПАТТЕРНЫ СОЧЕТАЕМОСТИ СВЯЗУЮЩИХ ЭЛЕМЕНТОВ СОЧИНИТЕЛЬНОЙ КОНСТРУКЦИИ В СОВРЕМЕННОМ ФРАНЦУЗСКОМ ЯЗЫКЕ

В статье представлены результаты статистической обработки сочетаемости многозначных сочинительных союзов во французском языке (*et* 'и', *ou* 'или', *mais* 'но') с конкретизаторами, постпозитивными элементами (частицами, наречиями), на основе анализа данных корпуса Frantext (с 2000 г. по н. в.) с применением статистических методов (t-score, Dice, MI). Выявлено 43 функционирующих конкретизатора, образующих 83 устойчивые комбинации с союзами. Установлено, что комбинаторный потенциал союзов различен: максимален у соединительного союза (40 комбинаций), минимален у разделительного (15 комбинаций). Продемонстрирована степень конкретизации союзов: союз *et* 'и' уточняется чаще всего (24,32 % случаев), союз *mais* 'но' – реже всего (6,69 %). Определены универсальные, специализированные и ведущие конкретизаторы многозначных союзов. На основании проведенного эксперимента по установлению корреляций в употребительности многозначных союзов и сочетаний союза с уточнителем в микро-диахронии сделано предположение о конъюнкционализации статистически устойчивых сочетаний, что требует проведения дополнительных исследовательских процедур.

Ключевые слова: союзы; конкретизаторы; сочинительные конструкции; сочетаемость; статистические паттерны; современный французский язык.

STATISTICAL PATTERNS OF CONNECTIVE ELEMENTS COMBINATION IN COORDINATION CONSTRUCTIONS IN CONTEMPORARY FRENCH

The article investigates the results of statistical processing of the compatibility of polysemic coordinating conjunctions in French (*et* 'and', *ou* 'or', *mais* 'but') with concretizers, postpositive elements (particles, adverbs) based on the analysis of data from the Frantext corpus (2000 to present), applying statistical methods (t-score, Dice, MI). 43 specifiers were identified, forming 83 stable combinations with the conjunctions. It was established that the combinatorial potential of the conjunctions varies: it is maximal for *et* (40 combinations) and minimal for *ou* (15 combinations). The analysis revealed that the frequency of conjunction specification varies: *et* is specified most often (24,32 % of cases), while *mais* is specified least often (6,69 %). Universal, specialized and leading concretizers of polysemic conjunctions are defined. Based on the conducted experiment to establish correlations in the usage of polysemic conjunctions and

combinations of a conjunction with a specifying element in microdiachrony, an assumption was made about the conjunctionalization of statistically stable combinations, which requires additional research procedures.

Key words: conjunctions; concertizers; coordination constructions; combination; statistical patterns; contemporary French.

Для выявления устойчивых соединений в сочинительной конструкции особую значимость приобретает применение корпусных технологий для анализа сочетаемости сочинительных союзов и их конкретизаторов. Статистические паттерны сочетаемости союзов и их конкретизаторов, вторых союзных элементов, демонстрируется на примере многозначных сочинительных союзов *et* 'и', *ou* 'или' и *mais* 'но' в современном французском языке на основании данных Национального корпуса Frantext с 2000 г. по настоящее время [1].

В результате анализа биграмм, содержащих союзы *et* 'и', *ou* 'или' и *mais* 'но', с последующей выдачей результатов, снабженных статистическими показателями об их устойчивости (t-score, Dice, MI), установлены 43 языковых единицы, регулярно функционирующих в роли их конкретизаторов. Количество зафиксированных в обозначенном сегменте корпуса Frantext коллокаций союзов и конкретизаторов как статистически устойчивых сочетаний слов [2] конечно и сводимо к 83 комбинациям (один из трех союзов и один из 43 конкретизаторов), что свидетельствует о наличии формальных и содержательных ограничений.

Для систематизации вторые союзные элементы в сочинительной конструкции на основании их семантических свойств обобщенно распределены на пять разных по количественному составу групп: противительные (*pourtant, du moins, au contraire, au moins, néanmoins, après tout, sinon, par contre, d'ailleurs, du reste, en outre, en revanche, au reste* – 13 ед.), каузальные (*donc, ainsi, en fait, de fait, en effet, par conséquent, à la vérité, effectivement, partant, conséquemment, c'est pourquoi* – 11 ед.), аддитивные (*aussi, encore, plutôt, de plus, encore que, de même (que), c'est à dire, au surplus, voire* – 9 ед.), темпоральные (*puis, cependant, toutefois, tantôt, sitôt, alors, enfin, ensuite, finalement* – 9 ед.), интенсифицирующие (*bien* – 1 ед.).

Комбинаторное разнообразие – максимальное количество комбинаторных вариантов (40) – фиксируется для союза *et* 'и' и его уточнителей с самой разнообразной семантикой: противительной (*Ma fille est un peu gâtée, mais elle pousse droit et au moins elle connaît papier* 'Моя дочь немного избалована, но растет хорошим человеком, и уж хотя бы с документами разбирается'¹), аддитивной (*Dans cette rue, où les siens ont vécu, où Ida a vécu, et aussi la mère de sa mère, et leurs ancêtres avant elles, Ethel décide qu'elle fera grandir ses enfants* 'На этой улице, где жила ее семья, где жила Ида, а также мать ее

¹ Здесь и далее примеры приводятся из корпуса [1]. Перевод наш – Ю.О.

матери и их предки до них, Этель решает, что она будет воспитывать своих детей'), темпоральной (*Je me sentais un peu comme dans ma montagne à regarder depuis les cimes les villages en fond de vallée. Et puis j'ai trouvé qu'on n'était toujours pas assez haut* 'Я словно нахожусь в горах, глядя с вершин на деревни в долине. А потом обнаружил, что мы всё еще недостаточно высоко'), каузальной (*Le sens du baptême est un engagement de l'être dans la foi, dans l'espoir de la résurrection. Il ne signifie rien pour un bébé, et par conséquent le baptême à la naissance n'a aucune valeur* 'Смысл крещения – это обязательство пребывать в вере и в надежде на воскресение. Для младенца оно ничего не значит, и поэтому крещение при рождении не имеет никакой ценности') семантикой.

Противительный союз *mais* 'но' конкретизируется 28 дополнительными постпозитивными элементами с противительной (*Peu à peu, ce projet était de moins en moins un geste masochiste, mais au contraire un geste joyeux, ludique, que je voulais faire partager* 'Постепенно этот проект все меньше походил на мазохистский жест, а наоборот, на радостный, игривый жест, которым мне хотелось поделиться'), аддитивной (*Je pleure. Je pleure pour moi, mais aussi pour tous les hommes et les femmes de mon pays que l'amour a quittés* 'Я плачу. Я плачу за себя, но также и за всех мужчин и женщин моей страны, которых покинула любовь'), каузальной (*Il connaissait des gens, et parmi les gens qu'il connaissait il se trouvait deux ou trois garçons qui sortaient de Fresnes et pouvaient faire ce genre de travail, mais alors il faudrait les rémunérer* 'Он знал людей, и среди них было два-три парня, приехавших из Френса и способных выполнять такую работу, но тогда им нужно было бы платить') семантикой.

Для разделительного союза *ou* 'или' характерно наименьшее количество допустимых комбинаций (15) с дополнительными элементами со схожей семантикой в постпозиции (*ou plutôt* 'или скорее', *ou alors* 'или тогда', *ou encore* 'или еще', *ou au contraire* 'или напротив', *ou sinon* 'или напротив' и т. п.).

Разная интенсивность проявления процесса конкретизации сочинительных союзов объективируется в относительных показателях: количество зафиксированных в корпусе сочетаний союза с его конкретизаторами относительно общего количества вхождений этого союза. Наибольшая степень конкретизации зафиксирована у союза *et* 'и', значение которого уточняется в 24,32 % случаях, или 13 332 от 54 802 вхождений союза. Значение союза *ou* 'или' конкретизируется введением дополнительных элементов в постпозиции к союзу в 11,33 % случаях, или 5322 от 46 960 вхождений союза. Значение союза *mais* 'но' по сравнению с *et* 'и' и *ou* 'или' подвержено этому процессу меньше всего: 6,69 % случаев, или 3364 от 50 294 вхождений союза.

Для иллюстрации сочетаемостного потенциала союзов и вторых союзных элементов в сочинительной конструкции приводится диаграмма на

рис. 1, отражающая как различную вовлеченность в процесс конкретизации языковых единиц в качестве второго союзного элемента в сочинительной конструкции с союзами *et* ‘и’, *ou* ‘или’ и *mais* ‘но’, так и способность вторых элементов либо сочетаться исключительно с одним сочинительным союзом, либо использоваться для конкретизации трех союзов.

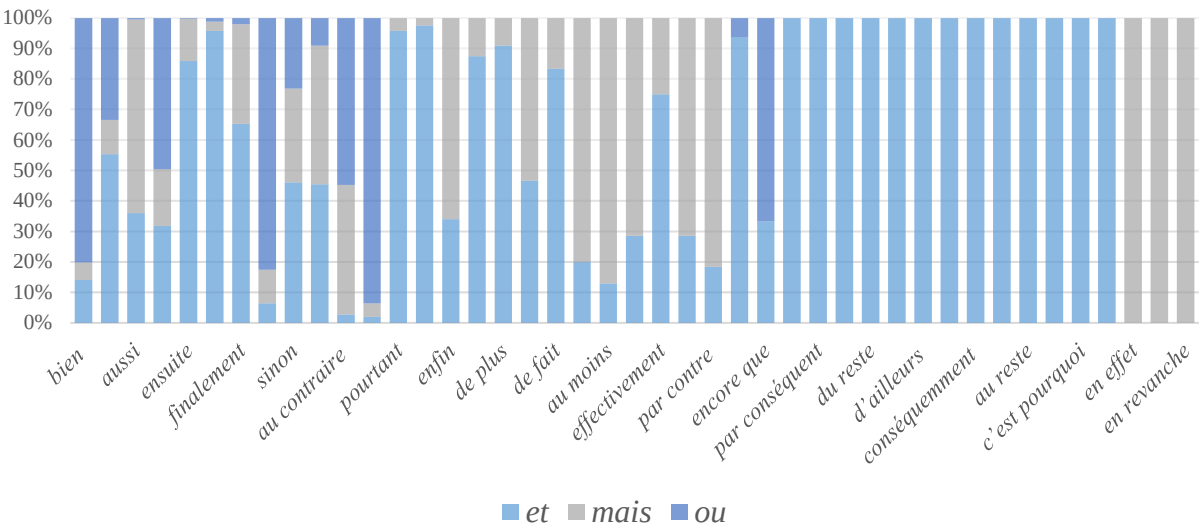


Рис. 1. Сочетаемостные паттерны французских сочинительных союзов *et* ‘и’, *ou* ‘или’ и *mais* ‘но’ и их конкретизаторов

В зависимости от сочетаемости с одним или более из трех многозначных союзов *et* ‘и’, *ou* ‘или’ и *mais* ‘но’ выделяются универсальные и специализированные конкретизаторы. Треть конкретизаторов является универсальными для многозначных союзов в современном французском языке, к ним относятся интенсифицирующее наречие *bien*, маркеры аддитивности *encore*, *plutôt*, *de même (que)*, темпорально-каузальные наречия *alors*, *finalement*, *ainsi*, а также единицы с противительной семантикой *sinon*, *du moins*, *au contraire*. Треть конкретизаторов специализируется исключительно для уточнения семантики союза *et* ‘и’ (темпоральное наречие *puis*, каузальные единицы *partant*, *conséquemment*, *par conséquent*, *c’est pourquoi*, противительные наречия *d’ailleurs*, *du reste*, аддитивные *c’est à dire*, *au surplus*, *voire*). Специализированными конкретизаторами союза *mais* ‘но’ являются наречные обороты *en revanche*, *en effet*, *à la vérité*. В сочетании с союзом *ou* ‘или’ статистически значимых специализированных конкретизаторов не выявлено.

Элементы с противительной семантики в сочетании с союзами *et* ‘и’ и *mais* ‘но’ преобладают: 12 и 10 из 13 возможных комбинаций для каждого союза. Элементы с аддитивной и темпоральной семантикой продуцируют одинаковое количество комбинаций с тремя союзами (9:5:5 из 9 возможных комбинаций для каждого союза), при этом для разделительного союза сочетания такого рода преобладают. Каузальные единицы сочетаются чаще с

соединительным союзом, чем с противительным или разделительным (9:7:1 из 11 возможных для каждого союза). В зафиксированных в современном французском языке комбинациях в сочинительной конструкции ее второй союзный элемент может проявлять коннективные свойства самостоятельно за исключением интенсифицирующего наречия *bien*, сочетающегося с тремя союзами.

Ведущую роль в процессе конкретизации того или иного союза выполняют разные элементы. Для союза *et* ‘и’ преимущественно используется наречие *puis*, сочетание с которым значительно превышает количество вхождений других сочетаний (54,43 % от 13 332). Для союзов *ou* ‘или’ и *mais* ‘но’ характерно преобладание в постпозиции к союзу наречий *bien* и *aussi* (48,55 % от 5 322 и 40,78 % от 3364 соответственно). Доля сочетаний других вторых союзных элементов в постпозиции к исследуемым союзам колеблется и составляет для союза *et* ‘и’ от 0,008 % до 7,8 %, для союза *ou* ‘или’ – от 0,019 % до 19,13 % и для союза *mais* ‘но’ – от 0,095 % до 13,83 %.

Использование в сочинительной конструкции конкретизаторов, которые в зависимости от контекстного окружения профилируют определенные логико-семантические отношения, может, на наш взгляд, сопровождаться падением частности того или иного многозначного союза, требующего уточнения. Так, в качестве эксперимента нами отобраны самые частотные элементы, а именно: сочинительный союз *et* ‘и’ и сочетание *et puis*. Для установления корреляций в их употребительности сопоставлено количество вхождений союза *et* ‘и’ и сочетания *et puis* на один миллион словоупотреблений (статистическая мера IPM). Замеры производились в микродиахронии (о дифференциации макро- и микродиахронии при статистической обработке корпусных данных высоко частотных единиц подробно в [3, с. 98–105]), т. е. на диахроническом отрезке современного французского языка с 1800 г. по настоящее время с десятилетним интервалом (см. рис. 2).



Рис. 2. Частотное распределение сочинительного союза **et** ‘и’ и союзного сочетания **et puis**

Если график построенной кривой частотного распределения сочинительного союза *et* ‘и’ в микродиахронии демонстрирует неоднократные скачки, то кривая частотного распределения союзного сочетания *et puis* на этом же диахроническом отрезке характеризуется сглаженностью, отсутствием резких колебаний, что объясняется несоразмерностью сопоставляемых количественных характеристик (значительное превышение употребительности союза *et* ‘и’). Сглаженность кривой частотного распределения сочетания не позволяет проследить тенденцию употребительности союзного сочетания *et puis*, поэтому отдельно выстраивается график частотного распределения союзного сочетания *et puis* в диапазоне его максимальных и минимальных показателей (см. рис. 3).

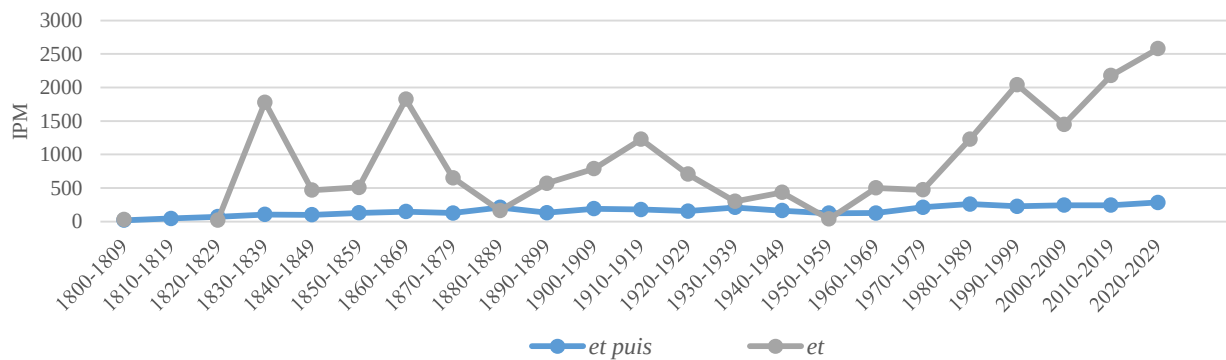


Рис. 3. Частотное распределение союзного сочетания **et puis**

Построенный график кривой частотного распределения союзного сочетания *et puis* отражает его растущую употребительность с множественными флуктуациями в определенные периоды. При сопоставлении двух кривых – кривой частотного распределения союза *et* ‘и’ (рис. 2) и кривой частотного распределения сочетания *et puis* (рис. 3) – обнаруживается синхронизация резкого падения употребительности союза *et* ‘и’ и скачка употребительности сочетания *et puis* на синхронных срезах 1840–1859 гг., 1880–1889 гг., 1930–1939 гг., 2000–2009 гг. Такого рода точки бифуркации на временной оси, представляющие собой пересечение разнонаправленных тенденций употребительности союза *et* ‘и’ и союзного сочетания *et puis*, свидетельствуют о компенсации семантической многозначности сочинительного союза *et* ‘и’ в ряде контекстов за счет его конкретизации наречием *puis*, которое имеет сложную семантическую структуру: выделяется темпоральное значение (временного следования) и значения, не относящиеся к темпоральной сфере (локативное, нумеративное, иерархическое и аддитивное) [4].

Исходя из положений теории грамматикализации (см. теорию грамматикализации [5]), полагаем, что на основании роста употребительности сочетания *et puis* допустимо его квалифицировать как грамматикализованный элемент. Вопрос наличия ряда существенных формальных и содержательных оснований (формальная нерасчлененность, линейное положение в структуре

высказывания, количество устойчивых соединений сочетания *et puis* с разного рода другими уточнителями, семантическая нерасчлененность и контекстная обусловленность значения) требует дополнительного изучения.

Таким образом, при общей тенденции к конкретизации семантики многозначных сочинительных союзов обнаруживается ряд частных ее проявлений во французском языке как по интенсивности самого процесса для каждого союза, так и по вовлеченности в него языковых единиц в качестве второго союзного элемента в сочинительной конструкции, что приводит к ограниченному образованию статистически значимых сочетаний. Перспективным видится установление формальных и содержательных ограничений на комбинаторные свойства союзных элементов в сочинительной конструкции.

ЛИТЕРАТУРА

1. Frantext. URL: <http://www.frantext.fr>. (дата обращения: 20.02.2025).
2. Хохлова М. В. Экспериментальная проверка методов выделения коллокаций // *Slavica Helsingiensia* 34. Инструментарий русистики: Корпусные подходы / Под.ред. А. Мустайоки и др. Хельсинки, 2008. С. 343–357.
3. Овсейчик Ю. В. Система единиц сочинения французского языка в диахронии. Минск : МГЛУ, 2023. 304 с.
4. Кобозева И. М. Контрастивный когнитивно-семантический анализ наречий временного следования *затем, потом* в русском и *puis, ensuite* во французском языке // *Lingue slave e lingue romanze a confronto: dalla frase al testo*. Medi Azioni. 36. 2023. P. A212–A232.
5. Lehmann C. *Thoughts on grammaticalization*. 3rd edition. Berlin : Language Science Press. 2015. 214 p.

УДК 811.511.111

Осьмак Наталья Андреевна

кандидат филологических наук,
заведующий кафедрой языков
Северной Европы
РГПУ им. А. И. Герцена
г. Санкт-Петербург, Россия

Natalia Osmak

PhD in Philology,
Chair of Northern European Languages
Department
Herzen State Pedagogical University
of Russia
St. Petersburg, Russia
Natalia.osmak@yandex.ru

Коваль Игорь Андреевич

магистрант
РГПУ им. А. И. Герцена
г. Санкт-Петербург, Россия

Igor Koval

Master's student
Herzen State Pedagogical University
of Russia
St. Petersburg, Russia
i@igorkoval.ru

КОРПУСНЫЕ ТЕХНОЛОГИИ В ИССЛЕДОВАНИИ И ПРЕПОДАВАНИИ ФИНСКОГО НЕВЕРБАЛЬНОГО ПОВЕДЕНИЯ¹

Актуальность исследования объясняется отсутствием специализированных корпусных инструментов для анализа финской невербальной коммуникации, в частности, ее кинетической составляющей. Несмотря на существование мультимодальных корпусов для других языков, аналогичный ресурс для финского языка пока не создан, что ограничивает как исследование финского коммуникативного материала, так и разработку учебных материалов, представляющих аспект невербального поведения. Целью статьи является презентация пилотного проекта мультимодального корпуса финского языка, включающего систему аннотации кинетических элементов и демонстрация его потенциала для исследовательской и педагогической деятельности. Методология включает в себя отбор репрезентативного материала (сериал «Luottomies»), его многоуровневую разметку в среде ELAN по авторской схеме с использованием функциональной классификации жестов, а также качественный и количественный анализ аннотированных данных. Результаты анализа свидетельствуют о достаточно умеренном использовании жестов в финской лингвокультуре с преобладанием жестов, дублирующих или сопровождающих вербальную информацию. Созданный корпус предлагает широкие возможности для дальнейшего исследования финской коммуникации, в том числе для сравнительного анализа межкультурной коммуникации и создания учебных материалов. К перспективам работы можно отнести расширение объема и жанрового разнообразия материала корпуса и совершенствование системы аннотации.

Ключевые слова: финский язык; невербальная коммуникация; мультимодальный корпус; жесты; корпусная лингвистика.

¹ Исследование выполнено за счет внутреннего гранта РГПУ им. А. И. Герцена (проект № 75-ВГ)

CORPUS TECHNOLOGIES FOR RESEARCHING AND TEACHING FINNISH NONVERBAL BEHAVIOR

The relevance of the study is explained by the absence of specialized corpus tools for analyzing Finnish nonverbal communication, particularly its kinetic component. Despite the existence of multimodal corpora for other languages, a comparable resource for the Finnish language has not yet been created. This limitation hinders both the research of Finnish communicative material and the creation of teaching materials representing the aspect of nonverbal behavior. The aim of the article is to present a pilot project of a multimodal corpus for the Finnish language, which includes a system for annotating kinetic elements, and to demonstrate its potential for research and pedagogical activities. The methodology involves selecting representative material (the series «Luottomies»), performing its multi-tier annotation in the ELAN environment using an original scheme based on a functional classification of gestures, and conducting qualitative and quantitative analysis of the annotated data. The results of the analysis indicate a relatively moderate use of gestures within Finnish communicative culture, with gestures replicating or accompanying verbal information being predominant. The created corpus offers broad opportunities for further research of Finnish communication, including comparative analysis of cross-cultural communication and the creation of teaching materials. Future perspectives for the work include expanding the volume and genre diversity of the corpus material and upgrading the annotation system.

Key words: Finnish language; nonverbal communication; multimodal corpus; gestures; corpus linguistics.

В настоящее время всё большее значение обретают лингвистические корпуса как для исследования разнообразных явлений и фактов языка, так и в качестве инструмента для преподавания различных дисциплин, сопряженных с языком и коммуникацией. Тем не менее, большая часть исследований в области корпусной лингвистики направлена на изучение именно языкового аспекта, в то время как комплексная сущность коммуникации также имеет потенциал быть исследованной и обработанной с применением корпусных технологий. Аналогичная ситуация обстоит и по части использования корпуса в образовательном процессе: в настоящее время преобладающим направлением в методике обучения иностранным языкам является использование корпуса как источника для создания учебных и лексикографических материалов, репрезентирующих реальную речевую и лингвокультурную ситуацию [1]. В контексте комплексного подхода к коммуникации актуализируется понятие мультимодальности, поскольку «Любое использование языка неизбежно является мультимодальным» [2, р. 386]. В данном поле встает вопрос об исследовании, а значит и дальнейшем преподавании других модальностей, в первую очередь, кинетической. Для подобных задач существуют и разрабатываются мультимодальные корпуса. В частности, можно отметить, например, созданный в рамках Национального Корпуса Русского Языка (НКРЯ) мультимедийный подкорпус (МУРКО), который используется в том числе для изучения кинетической составляющей русской коммуникации [3]. Данные исследования ценны, в первую очередь как инструмент изучения коммуникации с более объективной, числовой и количественной, точки зрения.

В то же время, на настоящий момент не существует корпусных инструментов, которые позволили бы изучать подобным образом финскую коммуникацию. Таким образом, актуальной является задача создания мультимодального корпуса для изучения финского коммуникативного поведения, в частности, с акцентом на кинетическую составляющую.

Задачи настоящей статьи состоят в:

1. описании предпосылок к созданию мультимодального корпуса для финской коммуникации;
2. описании разработанных принципов аннотации невербального поведения;
3. описании созданного на данный момент фрагмента мультимодального корпуса;
4. раскрытии потенциала корпуса в исследовательской и преподавательской деятельности.

Прежде, чем перейти к описанию корпуса, остановимся подробнее на понятии мультимодальности. Например, А. А. Кибрик считает, что мультимодальность, относится к разграничению поступающей информации по модальностям, то есть, в первую очередь, по органам восприятия человека с дальнейшим дроблением на каналы [4]. В видении Г. Е. Крейдлина мультимодальность состоит в «описании общих законов и правил взаимодействия в коммуникативном акте вербальных и невербальных знаков» [5, с 100]. В то же время, Г. Кресс считает мультимодальными явлениями взаимодействия между вербальными текстами и изображениями, видео, речью и жестами, размером и цветом текста [6]. Так, в его понимании в понятие мультимодальности включается мультимедийность [7]. По нашему мнению, мультимодальный корпус должен отвечать, в первую очередь, принципу нескольких модальностей, то есть строиться на материале, содержащем различные модальности, в первую очередь, вокальную и визуальную.

Следуя этому принципу, стоит обратить внимание на проекты, существующие для изучения финской коммуникации. В первую очередь, стоит обратить внимание на проект ‘Мультимодальный корпус туристических брошюр города Хельсинки’. Данный корпус представляет собой развороты туристических брошюр, текстовый и визуальный компонент. Стоит отметить, что данный корпус соответствует подходу к мультимодальности Г. Кресса, однако, не отвечает принципу нескольких модальностей, поскольку предоставляет материал исключительно визуально. Кроме того, язык материала данного корпуса – английский, что делает данный проект нерелевантным в изучении финской коммуникации [8]. Стоит отметить, что для финского коммуникативного поведения существуют проекты, направленные на изучение коммуникации с учетом нескольких модальностей, в частности, кинетического канала визуальной модальности [9]. Однако, подобные проекты не используют корпусные технологии, что подчеркивает актуальность создания мультимодального корпуса для финской коммуникации.

Одной из важнейших задач в создании подобного корпуса является разработка систем аннотации невербальной коммуникации. Среди средств невербальной коммуникации можно выделить несколько аспектов как, например, проксемика, рассматривающая коммуникативную роль пространства, такесика или гаптика, изучающая тактильную коммуникацию и кинесика, изучающая «движения тела в связи с невербальными аспектами межличностного общения» [10, р 9].

В рамках настоящей статьи акцент сделан на невербальную составляющую коммуникации и корпусные инструменты её анализа. Данный выбор обусловлен тесным и активным взаимодействием информации, содержащейся в кинетических средствах коммуникации с вербальной составляющей при непосредственном акте коммуникации, в то время как средства невербальной коммуникации в рамках проксемики и такесики часто характеризуют культуру в целом и дают некоторый объем информации, неизменный при любой коммуникации. Помимо этого, кинетический компонент представляется наиболее подходящим для анализа с использованием корпусных технологий, поскольку при изучении акта коммуникации возможно выделить отдельные элементы для аннотации и анализа.

В качестве материала для создания корпуса был выбран сериал «*Luottomies*» «Надёжный чувак» режиссеров Лаури Нурксе, Ялмари Хеландера и Кари Кетонена в жанре бытовой комедии. Выбор обусловлен несколькими факторами:

- Жанр бытовой комедии позволяет наблюдать, как архетипичные образы финнов ведут себя как в привычной бытовой обстановке, так и в нетипичных ситуациях, стимулирующих проявление различных эмоций и состояний.
- Художественное произведение, созданное представителями финской коммуникативной среды для таких же потребителей, дает возможность утверждать о достаточной достоверности поведения персонажей в сравнении с реальными ситуациями коммуникации.
- В условиях ограниченной доступности к сотрудничеству с финскими респондентами данный материал, тем не менее, дает возможность осуществления проекта.
- Кинетические средства невербальной коммуникации не могут быть полностью искусственными даже в условиях кинопроизводства.

Для пилотного корпуса был отобран 1 сезон, состоящий из 8 серий, общей длительностью 69 минут, который был обработан в свободно распространяемом программном обеспечении ELAN[11], поскольку оно обеспечивает полноценный и удобный инструментарий для работы с видеоматериалом, как на этапе обработки, так и на этапе представления и работы с обработанным материалом.

Таким образом, был разработан и реализован пилотный проект мультимодального корпуса и аннотации невербальных кинетических средств для финской коммуникации.

За основу разработанной системы разметки была взята функциональная классификация Г. Е. Крейдлина [12, с 61–63]. Так, на основании функций по отношению к вербальному компоненту были выделены следующие типы элементов:

1. элемент дублер (функция дублирования вербальной информации)
2. элемент-противопоставитель (функция противоречия вербальной информации)
3. элемент-заменитель (функция замещения вербальной информации)
4. элемент-усилитель (функция усиления вербальной информации)
5. элемент-регулятор (функция регулирования процесса коммуникации)
6. элемент-сопроводитель (не вступает во взаимодействие с вербальной информацией) (данный элемент не сопоставляется с функциями, выделяемыми Г. Е. Крейдлиным, добавлен на основании противопоставления наличия функции по отношению к вербальному компоненту и ее отсутствия).

Также, аннотация включает в себя локализацию органа-исполнителя (рука, голова, лицо и т. п.), описание элемента, то есть описание движения и/или знака.

Помимо этого, аннотация включает в себя метку, относящуюся к информации, которую элемент невербальной коммуникации несет самостоятельно, независимо от вербального компонента. Данная метка присваивается только элементам, для которых справедливо наличие информации, не выраженной вербально, а также элементам, которые не сопровождают вербальное высказывание. Были выделены следующие две метки:

1. [инф] – информация фактологического характера (направление, количество и т.п.)

2. [эмо] – Информация об эмоции, состоянии (раздражение, гнев и т. п.)

Таким образом, в среде ELAN было выделено 5 слоев аннотации:

1. Gest – локализация органа-исполнителя и описание движения, знака;
2. Gest type – указание типа элемента невербальной коммуникации (усилитель, сопроводитель и т.п.);
3. Gest index – метка об информации, не выраженной вербально – [инф],[эмо];
4. Gest mean – расшифровка значения элемента невербальной коммуникации (указание, недоумение и т.п.);
5. Text – текстовая расшифровка вербального компонента, соответствующего выделенному невербальному.

Мультимодальный корпус в среде ELAN обеспечивает ряд возможностей для исследования размеченного материала с применением корпусных технологий:

- Навигация по временной линии видеоматериала;
- Поиск по любым вхождениям в материале;
- Поиск по слоям разметки:
 - тип элемента,
 - орган-исполнитель,

- значение элемента,
- наличие и тип информации, не выраженной вербально;
- Поиск по словоформам;
- Количественная оценка вхождений.

По результатам исследования финского невербального поведения на обработанном материале с применением вышеназванных инструментов можно привести ряд результатов.

Так, на всей продолжительности был выделен 81 элемент невербальной коммуникации, типы которых в количественном соотношении распределились следующим образом (рис.1):

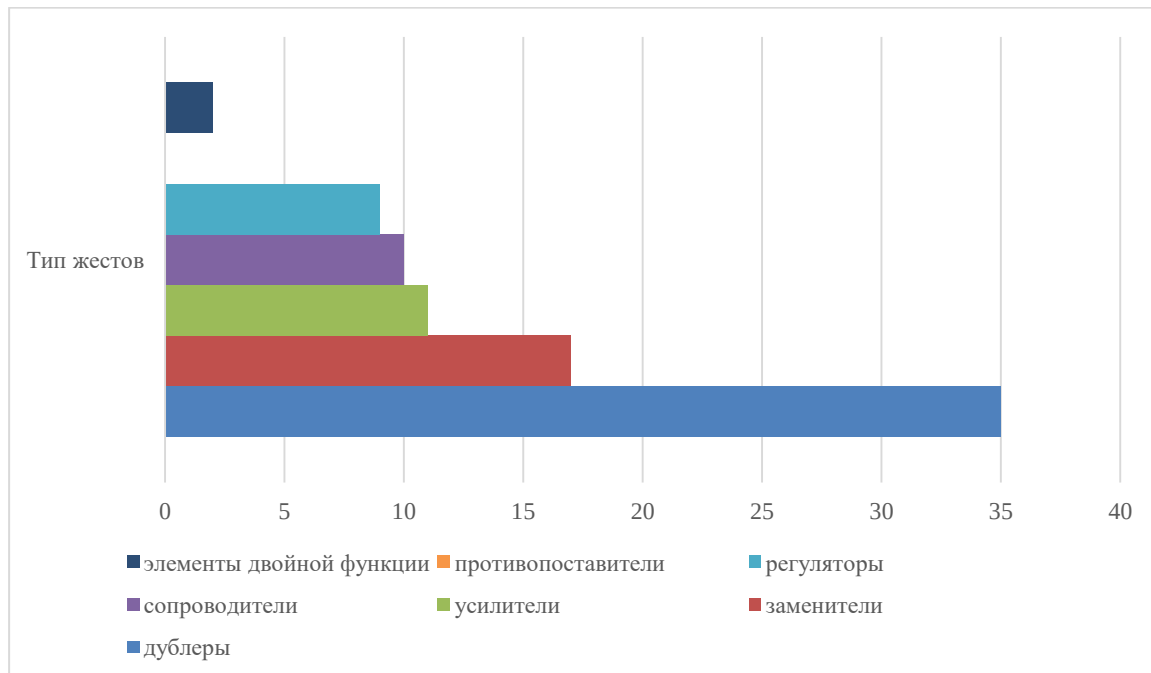


Рис. 1. Количественное соотношение типов жестов в пилотном материале

Элементы локализованы по органам-исполнителям следующим образом (рис. 2):

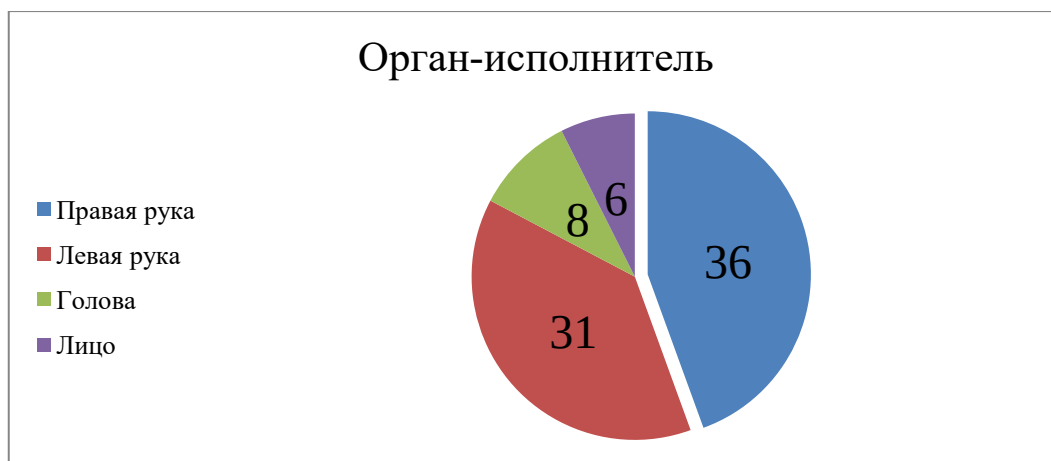


Рис. 2. Количественное соотношение органов-исполнителей жестов в пилотном материале

При этом 38 элементов несут информацию, которая не выражена вербально. Из них 18 элементов несут информацию фактологического характера, 21 – информацию об эмоциях, состоянии, 1 элемент – оба вида информации.

На основании первичного анализа полученных данных можно сделать следующие выводы:

1. Для финского коммуникативного поведения, представленного в изученном материале, характерно умеренное использование элементов невербальной коммуникации

2. Большая часть элементов невербальной коммуникации не несет информацию самостоятельно, а лишь дублирует и дополняет информацию, выраженную вербально

3. Наиболее распространены элементы указания и иллюстрации, выполненные одной (правой) рукой.

На основании приведенных данных и результатов исследования можно говорить о широком исследовательском потенциале корпуса и разработанной системы разметки.

Помимо потенциала для исследования коммуникативного поведения, данный корпус обладает, также, потенциалом для преподавания. Помимо того, что данные мультимодального корпуса могут быть использованы в качестве иллюстративного, аудиовизуального материала, а также источника конкретных примеров для отбора учебного материала в преподавании непосредственно финского языка, функциональные слои аннотированных невербальных средств коммуникации позволяют применять корпус в преподавании дисциплин, связанных с межкультурной коммуникацией. Возможности корпуса обеспечивают удобный инструментарий для, например, заданий на сравнение коммуникативного поведения в различных контекстах, а также в различных культурах с применением схожих продуктов для таких культур. Помимо этого, возможно использование корпуса для заданий и упражнений на наблюдение и описание финской коммуникативной практики.

Таким образом, созданный проект мультимодального корпуса отвечает актуальным проблемам области и обеспечивает возможности как в исследовании, так и в преподавании финского коммуникативного поведения, разработанная и примененная система аннотации кинетических средств невербальной коммуникации позволяет изучать финскую коммуникацию более комплексно, учитывая кинетическую невербальную составляющую. Также, представленный продукт имеет потенциал к развитию, расширению и совершенствованию. В первую очередь, расширение объема, жанра и типа исследуемого материала позволит составить более полную картину финского коммуникативного поведения. Кроме того, система разметки также может быть усовершенствована в ходе дальнейшей работы путем добавление большего количества слоев аннотации, основанных на других существующих классификациях средств невербальной коммуникации.

ЛИТЕРАТУРА

1. Павлова О. Ю. Использование языковых корпусов в обучении иностранному языку // Язык и культура. 2021. № 54. С. 283–298. DOI: 10.17223/19996195/54/16.
2. Scollon R. Multimodality and the language of politics Encyclopedia of language and linguistics, 2nd ed. / Eds. Yu. I. Alexandrov et al. Oxford: Elsevier, 2006. V. 9. P. 386–387.
3. Колмогорова А. В., Мигаль А. В., Сергеева М. О. Количественный анализ жестов в советском кино на материале Мультимедийного корпуса Национального корпуса русского языка // Культура и текст: сетевое издание. 2024. № 2 (57). С. 181–191.
4. Кибрик А. А. Мультимодальная лингвистика // Когнитивные исследования: сборник научных трудов / под редакцией Ю. И. Александрова, В. Д. Соловьева. Москва: Институт психологии РАН, 2010. Вып. 4. С. 135–152.
5. Крейдлин Г. Е. Семиотическая концептуализация тела и проблема мультимодальности // Экология языка и коммуникативная практика. 2014. № 2. С. 100–120.
6. Kress G. What Is Mode? // A Handbook of Multimodal Analysis / ed. by C. Jewitt. L. – N. Y.: Routledge, 2009. P. 54–67.
7. Kress G., van Leeuwen T. Multimodal Discourse: The Modes and Media of Contemporary Communication. – L.: Arnold, 2001. – vii, 142 p.
8. Hiippala T. Helsingin kaupungin matkailuesitteiden multimodaalinen korpus // Terra – 2016 – № 128 (2). P. 75–85.
9. Paananen J. Kuinka eleet helpottavat yhteisymmärrystä? Ikoniset ja deiktiset eleet monikulttuurisilla lääkärin vastaanotoilla. // Puhe ja kieli. 2015. № 35 (2). P. 73–96.
10. Birdwhistell R. L. Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture. – Washington, DC: Department of State, Foreign Service Institute, 1952. 75 p.
11. ELAN: a Professional Framework for Multimodality Research / P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, H. Sloetjes // Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation, 2006. URL: <https://archive.mpi.nl/tla/elan>.
12. Крейдлин Г. Е. Невербальная семиотика: Язык тела и естественный язык. М. : Новое литературное обозрение, 2002. 592 с.

УДК 811.581'373.612.2+003.212 (045)

Серова Ангелина Сергеевна
преподаватель кафедры
теории и практики китайского языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Angelina Serova
Lecturer of the Department of
Theory and Practice of the Chinese
Language
Belarusian State University
of Foreign Languages
Minsk, Belarus
Angelina_Serova@hotmail.com

КОМБИНАТОРИКА ЗНАКА 给 ‘ДАВАТЬ’ С ИЕРОГЛИФАМИ СЕМАНТИЧЕСКОГО ТИПА «СВЯЗИ И ОТНОШЕНИЯ»: ОПЫТ КОРПУСНОГО ИССЛЕДОВАНИЯ

В данной статье исследуется корреляция между семантической сочетаемостью иероглифического знака 给 ‘давать’ и его структурной организацией в метафорических конструкциях. На основе анализа выявляются и описываются ключевые макро- и микро-структурные модели их формирования.

Ключевые слова: метафорическое сочетание; иероглифический знак 给 ‘давать’; макроструктура; микроструктура; модель; семантический тип; китайский язык.

COMBINATORICS OF THE SIGN 给 ‘TO GIVE’ WITH HIEROGLYPHS OF THE SEMANTIC TYPE ‘CONNECTIONS AND RELATIONS’: A CORPUS-BASED STUDY

This article examines the correlation between the semantic compatibility of the Chinese character 给 ‘to give’ and its structural organization within metaphorical phrases. Through analysis, the key macro- and microstructural patterns of their formation are identified and described.

Key words: metaphorical phrase; Chinese character 给 (‘to give’); macrostructure; microstructure; model; semantic type; Chinese language.

Несмотря на многочисленные походы к изучению метафоры наивысший пик популярности исследований, посвященных метафоре, произошел после публикации работы Дж. Лакоффа и М. Джонсона «Метафоры, которыми мы живем» [1]. Мысль Дж. Лакоффа о том, что метафора есть нечто большее чем языковое явление, поставила перед лингвистами новую задачу: переосмыслить метафорические образные сочетания через призму мыслительных процессов человека. В действительности мы можем утверждать, что метафоры не только позволяют нам познавать новое с помощью когнитивных переносов, но и во многом формируют наше мышление, создавая особые паттерны восприятия окружающей действительности. Учитывая, что эта

действительность разнится от народа к народу, можно утверждать, что метафора в том числе участвует в формировании культурного кода и самоидентичности человека. Поэтому изучение когнитивных метафор не просто углубляет знания о структуре и специфике конкретного языка, но и вскрывает глубинные модели мышления его носителей, что особенно актуально в условиях непрерывной глобализации и активного внедрения технологии искусственного интеллекта во все сферы жизни человека.

Несмотря на множество исследований в области когнитивной метафоры, область сочетаемости до сих пор остается наименее изученной в современной синологии. В связи с чем нами была совершена попытка подробнее изучить структурные и семантические особенности образных сочетаний в китайском языке. Для данной цели нами был выбран процессуальный знак 给 ‘давать’, т. к. процессы передачи объектов лежат в основе человеческого бытия, а данный знак является одним из наиболее ярких представителей реализации этих процессов в языке.

Методом сплошной выборки из Национального корпуса китайского языка BCC 语料库 (BLCU Corpus Center) [2] нами было отобрано 312 метафорических сочетаний типа 给个理由 ‘дать причину’, 给他机会 ‘дать ему шанс’, 给不了爱情 ‘не мочь дать любви’, 给个答复 ‘дать ответ’, 给世界礼物 ‘подарить миру подарок’, 给一个笑容 ‘подарить улыбку’, 给个五星 ‘оценить в пять звезд’. На основании семантического словаря Н. Ю. Шведовой [3] знаки при процессуальном знаке 给 ‘давать’ были отнесены к одиннадцати возможным семантическим типам, вторым по продуктивности из которых является семантический тип «Связи, отношения. Содержание и форма. Структура. Метод» (22,8 %, 72 лексические единицы). Данный тип представлен шестью подтипами: «Связи. Отношения» (7,69 %, например, 给个鼓励 ‘оказать поддержку’, 给个依靠 ‘дать опору’, 给个婚礼 ‘устроить свадьбу’, 给个家 ‘создать семью»), «Модальность» (1,32 %, например, 给个真相 ‘дать правду’, 给人民批评的自由 ‘дать народу свободу критики»), «Ценность, её критерии» (5,45%, например, 给个友情价 ‘предложить дружескую цену’, 给经济损失 ‘нанести ущерб экономике»), «Обусловленность» (2,88 %, например, 给上司做的道理 ‘дать начальству причину’, 给个目标 ‘дать задачу’, 上级部门在给下面动力 ‘начальство дает мотивацию подчиненным»), «Содержание и форма. Структура. Метод» (4,49 %, например, 给个人的做法 ‘дать индивидуальный подход’, 给一条出路 ‘дать выход [из ситуации]’, 给个政策 ‘предложить политическую стратегию»), «Социальное взаимодействие в виртуальной среде» (0,96 %, например, 给个邀请码 ‘дать инвайт-код’, 给个微信 ‘дать вичат’).

Метафорические сочетания с процессуальным знаком 给 ‘давать’ и лексемами представленного семантического типа были проанализированы

с точки зрения их структурных особенностей, в результате чего было выделено 7 макроструктур (состоящих и более частных моделей – микро-структур) по которым могут образовываться подобные сочетания.

Структурные модели метафорических сочетаний, построенных по схеме «иероглифический знак 给 ‘давать’ + знаки, семантического типа «Связи, отношения. Содержание и форма. Структура. Метод»»¹

№ п/п	Структурные модели метафорических сочетаний, построенных по модели «иероглифический знак 给 ‘давать’ + знаки, семантического типа “Связи, отношения. Содержание и форма. Структура. Метод”»	Количество, %
1	给 + 个 + O	36,1
1.1	给 + 个 + O	30,55
1.2	给 + 个 + Mo的 + O	2,77
1.3	给 + 个 + Mo + O	2,77
2	给 + Or / Os + O	27,77
2.1	给 + (Mv) + Or + (Mo的) + O	20,83
2.2	(Mv) + 给 + Or + C + L + (Mo 的) + O	6,94
3	给 + O	11,11
3.1	(Mv) + 给 + (C + L) + (Mo的) + O	9,72
3.2	给 + Mv + O	1,38
4	Sr / Ss + 给 + Or / Os + O	8,33
4.1	Sr + (Mv) + 给 + (C + L) + Or + (Mo) + O	6,94
4.2	Sr + (Mv) + 给 Mv + Or + Mo的 + O	1,38

¹ Условные обозначения:

Sr – субъект действия, обозначающий
живых существ
Ss – субъект действия, обозначающий
неживых существ
A – косвенный объект действия (адресат)
Or – одушевленный косвенный объект
Os – неодушевленный косвенный объект
O – прямой объект действия

Mv – процессуальный
модификатор
Mo – объектный модификатор
C – числительное
L – счетное слово
() – вариативность

5	给与	6,94
5.1	(Sr / Ss) + Mv + 给与 + (Or) + (Mo 的) + O	5,55
5.2	Sr + 给与 + 了 + (Mo的) + O	1,38
6	Sr / Ss + 给 + O	5,55
6.1	Sr /Ss + (Mv) + 给 + C + L + (Mo 的) + O	4,16
6.2	Sr / Ss + (Mv) + 给 + O	1,38
7	O + (Sr/Ss) + 给 + (Or/Os)	4,16
7.1	连 + (Mo的) + O + (Mv) + 给 + (Mv)	2,77
7.2	O + Sr + Mv + 给	1,38

Наиболее продуктивной макроструктурой (36,1 %) построения образных сочетаний с процессуальным знаком 给 ‘давать’ и единицами семантического типа «Связи, отношения. Содержание и форма. Структура. Метод» является 给 + 个 + O. Особенностью данной структуры является наличие связующего счетного слова 个, придающего сочетанию разговорный характер. Этим объясняется ее большая частотность в семантических подтипах «Связи, отношения» (11,1 %) и «Содержание и форма. Структура. Метод» (9,7 %), т. к. лексемы данных подтипов в том числе характерны для разговорного стиля общения: 给个鼓励 ‘оказать поддержку’, 给个肩膀 ‘дать плечо поддержки’, 给个恩典 ‘оказать милость’, 给个解决办法 ‘дать решение [проблемы]’, 给个台阶下 ‘дать возможность выйти из неловкой ситуации’. Значима данная макроструктура также для образования сочетаний с единицами семантического подтипа «Социальное взаимодействие в виртуальной среде», характерными для сферы интернет-коммуникации, где счетное слово наделяет высказывание экспрессивной окраской непринужденности 给个支付宝 ‘пришли-ка [свой] алипэй’, 给个邀请码 ‘поделись-ка инвайт-кодом’, 给个微信 ‘дай-ка [свой] вичат’. Макроструктура 给 + 个 + O включает 3 микроструктуры: 1) 给 + 个 + O (30,55 %), 2) 给 + 个 + Mo的 + O (2,77 %), 3) 给 + 个 + Mo + O (2,77 %). Последние две микроструктуры характеризуются наличием уточняющего объектного модификатора при прямом объекте, например, 给个人的做法 ‘дать индивидуальный подход’, 给个不一样的成绩 ‘показать разные результаты’, 给个最佳撒 ‘дать самое лучшее’, 给个解决办法 ‘предложить решение’. Равный процент структур с атрибутивной частицей 的 и без нее говорит о ее вариативности.

Второй по частотности макроструктурой является 给 + Or / Os + O (27,77 %), в которой помимо прямого объекта процессуальный знак 给 ‘давать’ сочетается с одушевленным или неодушевленным косвенным объек-

том действия. Процессуальный знак в данной макроструктуре может оформляться процессуальным модификатором, позиционирующимся как до, так и после процессуального знака. Прямой объект действия может распространяться объектным модификатором с атрибутивной частицей 的 (как в микроструктуре 给 + (Mv) + Or + (Mo的) + O (20,83 %)), а также числительно-счетной группой, конкретизирующей количество объекта (как в микроструктуре (Mv) + 给 + Or + C + L + (Mo 的) + O (6,94 %)). Наиболее часто прямой объект в сочетаниях, образованных по данной макроструктуре, представлен лексемами семантического подтипа «Ценностность, ее критерии» (11,1 %), как в примерах: 给世人一个样板 ‘подать пример всему миру’, 给中国消费者最大的尊重 ‘оказать максимальное уважение китайским потребителям’, 给经济损失 ‘нанести ущерб экономике’. Такую тенденцию можно объяснить тем, что единицы с ценностной семантикой обычно нуждаются в косвенном объекте действия, на который направлено их воздействие. Нуждаются в косвенном субъекте и лексемы подтипа «Связи. Отношения» (6,91 %), например, 给客人冷漠对待 ‘проявить равнодушие к гостям’, 给不了她应有的回报 ‘не мочь дать ей заслуженного вознаграждения’, 给了我孤独感 ‘принесло мне чувство одиночества’. Лексемы других семантических подтипов также могут образовывать метафорические сочетания по данным микроструктурам, но реже: 给别人一条活路 ‘оставить другим путь к выживанию’, 给失败者一分额外的收获 ‘дать неудачникам дополнительный шанс’ («Содержание и форма. Структура. Метод», 4,16 %), 要给人民批评的自由 ‘необходимо дать народу свободу критики’ («Модальность», 2,8 %), 给你足够的原因 ‘дать тебе достаточно обоснований’ («Обусловленность», 2,8 %).

В 11,11 % случаев образные сочетания, построенные по модели «иероглифический знак 给 ‘давать’ + знаки, семантического типа «Связи, отношения. Содержание и форма. Структура. Метод» образуются по структуре 给 + O. Данная макроструктура характеризуется отсутствием субъекта действия или косвенного объекта действия, что смещает смысловой акцент на сам процесс метафорической передачи абстрактного прямого объекта действия. Такая особенность чаще всего проявляется в сочетаниях с лексемами подтипа «Содержание и форма. Структура. Метод» (5,55 %), например, 给两条路 ‘предложить два пути’, 只给一个话题 ‘лишь выдвинуть одну тему для разговора’, т. к. в подобных метафорах субъект и адресат действия вторичны. Процессуальный знак обычно распространяется процессуальным модификатором, локализуемым до него (как в микроструктуре (Mv) + 给 + (C + L) + (Mo的) + O (9,72 %)), объект действия может вариативно оформляться объектным модификатором с атрибутивной частицей 的 и/или числительно-счетной группой, конкретизирующей его количество. Например, 多给一些支持 ‘чаще оказывать поддержку’, 不给一定松动的余地 ‘не оставить пространства для маневра’, 给一条出路 ‘предоставить выход [из ситуации]’, 给一路输

出的方法 ‘предложить способ экспорта’. Процессуальный модификатор может также локализоваться после процессуального знака (как в микро-структуре 给 + Mv + O), однако подобные сочетания встречаются реже, лишь в 1,38 % случаев. Прямой объект при этом не распространен: 给不了尊严 ‘не в силах сохранить честь’.

Частотность сочетаний, строящихся по модели, в которой задействованы все участники коммуникативного акта, а именно, субъект, косвенный объект и прямой объект действия (Sr + 给 + Or / Os + O), составляет 8,33 %. В 6,9 % случаев сочетания образуются по микроструктуре Sr + (Mv) + 给 + (C + L) + Or + (Mo) + O, где косвенный объект действия может конкретизироваться числительно-счетной группой, прямой объект распространяться объектным модификатором, а сам процессуальный знак – процессуальным модификатором. Например, 上级部门在给下面压力 ‘вышестоящие органы оказывают давление на низшие уровни’, 你给下属实惠 ‘ты должен обеспечить подчиненным материальную выгоду’, 世界各国友好人士给中国人民热情援助 ‘дружественные представители всех стран мира оказали китайскому народу горячую помощь’. Единичной (1,38 %) является микроструктура Sr + (Mv) + 给 Mv + Or + Mo的 + O, в которой процессуальный знак оформляется сразу двумя процессуальными модификаторами, а прямой объект объектным модификатором с атрибутивной частицей 的: 他又给不了你什么实际的好处 ‘он опять не может предоставить тебе никакой реальной выгоды’. Сочетания, образуемые по макроструктуре (Sr + 给 + Or / Os + O) равнозначно образуются с лексемами трех семантических подтипов, а именно «Связи. Отношения» (2,77 %), «Ценностность, её критерии» (2,77 %), «Обусловленность» (2,77 %).

Метафорические сочетания с лексемами семантического типа «Связи, отношения. Содержание и форма. Структура. Метод» также могут встречаться в официальных контекстах. Об этом свидетельствует возможность их образования по микроструктурам (Sr / Ss) + Mv + 给与 + (Or) + (Mo 的) + O (5,55 %) и Sr + 给与 + 了 + (Mo的) + O (1,38 %), в которых представлен двусложный вариант изучаемого процессуального знака 给与 ‘давать’, характерный для официально-делового и публицистического стилей речи. Лексемы только двух семантических подтипов («Связи, отношения» и «Ценностность, её критерии») могут участвовать в образовании таких сочетаний, например, 给与非洲儿童更多的关怀和保护 ‘проявлять больше заботы и обеспечивать защиту африканским детям’, 给与更多的理解和支持 ‘оказывать больше понимания и поддержки’, 给与中标企业合理的利润 ‘обеспечивать компаниям-победителям тендеров разумную прибыль’, 给与优待 ‘предоставлять льготы’. Структурно представленные сочетания характеризуются вариативным наличием субъекта и косвенного объекта действия, процессуальный знак обычно оформляется процессуальным модификатором, прямой объект действия мо-

жет распространяться объектным модификатором с атрибутивной частицей 的. В редких случаях после процессуального знака может употребляться частица 了, указывающая на завершенность действия: 给与了不小的影响 ‘оказало немалое влияние’.

Косвенный субъект действия в метафорических сочетаниях с процессуальным знаком 给 ‘давать’ и единицами семантического типа «Связи, отношения. Содержание и форма. Структура. Метод» опускается в 5,55 % случаев (макроструктура Sr / Ss + 给 + O). Смысловой акцент смещается на субъект действия, а высказывание приобретает более общий, неконкретный характер, например, 国家会给一些扶持 ‘государство окажет определенную поддержку’, 时间会给一个公正的结果 ‘время даст справедливый результат’. Как видно из примеров процессуальный знак в подобных сочетаниях может оформляться процессуальным модификатором. Прямой объект, представленный лексемами семантического подтипа «Связи. Отношения», требует оформление счетным словом с числительным, а также вариативно может распространяться уточняющим объектным модификатором с атрибутивной частицей 的, образуя микроструктуру Sr / Ss + (Mv) + 给 + C + L + (Mo 的) + O (4,16 %). Единичной является микроструктура Sr / Ss + (Mv) + 给 + O (1,38 %) в которой прямой объект ничем не оформлен и семантически относится к подтипу «Обусловленность»: 双方都给让步 ‘обе стороны пошли на уступки’.

Наименее частотной макроструктурой образования метафорических сочетаний с процессуальным знаком 给 ‘давать’ и единицами семантического типа «Связи, отношения. Содержание и форма. Структура. Метод» является O + (Sr/Ss) + 给 + (Or/Os) (4,16 %), в которой за счет инверсивного порядка предложения с вынесенным прямым объектом создается смещение логического центра предложения на тему высказывания. Характерная для разговорного стиля речи макроструктура O + (Sr/Ss) + 给 + (Or/Os) представлена двумя микроструктурами: 1) 连 + (Mo 的) + O + (Mv) + 给 + (Mv) (2,77 %), 2) O + Sr + Mv + 给 (1,38 %). В наиболее встречающейся микроструктуре инверсия создается с помощью союза 连: 连让医生开口的机会都不给 ‘даже возможности дать врачу высказаться не предоставили’, 连起码的婚姻都给不了 ‘даже минимально брака не может дать’. Вынесенный прямой объект в данных примерах распространяется объектным модификатором, а процессуальный знак – процессуальным модификатором. Реже прямой объект выносится вперед без вспомогательных элементов, однако может выделяться пунктуационно в письме и паузой в речи: 帮助, 他一点也没有给 ‘помощи, от него даже минимальной не дождешься’.

Проведенный анализ показал, что лексемы семантической группы «Связи, отношения. Содержание и форма. Структура. Метод» образуют с процессуальным знаком 给 ‘давать’ широкий спектр метафорических соче-

таний характерных как для разговорных, так и для формально-официальных стилей общения. Несмотря на то, что один и тот же подтип семантической группы «Связи, отношения. Содержание и форма. Структура. Метод» может участвовать в построении разных моделей, тем не менее в каждой из макроструктур можно выделить преобладающий семантический подтип прямого объекта. Исходя из чего можно сделать вывод, что лексико-семантический тип прямого объекта влияет на структурные особенности метафорических сочетаний.

ЛИТЕРАТУРА

1. Лакофф Д. Метафоры, которыми мы живем / Дж. Лакофф, М. Джонсон. М. : УРСС, 2004. 252 с.
2. BCC语料库 (BLCU Corpus Center). URL: <https://bcc.blcu.edu.cn/lang/zh>. (дата обращения: 08.05.2025).
3. Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений / РАН. Ин-т рус. яз.; под общей ред. Н. Ю. Шведовой. М. : Азбуковник, 2003.

УДК 811.11

Яковлева Виктория Вадимовна
кандидат филологических наук,
доцент кафедры иностранных языков
Национальный исследовательский
университет «МЭИ»
г. Москва, Россия

Viktoria Iakovleva
PhD in Philology,
Associate Professor of the Department
of Foreign Languages
National Research University «Moscow
Power Engineering Institute»
Moscow, Russia
iakovlevavv@mail.ru

ЛИНГВОСЕМАНТИЧЕСКАЯ СТРУКТУРИЗАЦИЯ ГЛАГОЛОВ *SCHWIMMEN* И *ПЛЫТЬ* В НЕМЕЦКОМ И РУССКОМ ЯЗЫКАХ

Статья посвящена сопоставительному анализу мотивационных структур немецких и русских глаголов движения на примере формантов от основ *schwimmen* и *плыть*. Цель исследования – выявление системных связей между приставочными глаголами двух языков с учетом их семантики и направленности движения. Актуальность работы обусловлена необходимостью изучения морфо-семантической асимметрии и степени грамматикализации пространственных отношений в немецком и русском языках. Методология включает корпусный анализ (DWDS, DGD, словарь С. И. Ожегова), компонентное разложение семантики и контрастивный подход. В немецком языке выявлено 18 приставочных глаголов от *schwimmen*, которые классифицированы по семантическим кластерам. Сравнительный анализ выявил различия в способах кодирования пространственных значений: немецкие префиксы передают более точную информацию, тогда как русские глаголы чаще требуют контекстуального уточнения.

Установлено, что немецкий язык обладает большей степенью грамматикализации пространственных отношений, в то время как русский компенсирует это синкретизмом приставок и гибкостью контекста. Результаты представлены в сопоставительной таблице, иллюстрирующей лакунарность некоторых соответствий. Работа вносит вклад в теорию языковой концептуализации движения и может быть полезна для переводоведения и лингводидактики. Перспективой исследования обозначен анализ частотности употребления глаголов в диахроническом аспекте.

Ключевые слова: глаголы движения; префиксы; корпус немецкого языка; лакулярность; лингвотопология; контрастивная лингвистика.

LINGUOSEMANTIC STRUCTURING OF MOTION VERBS *SCHWIMMEN* AND *ПЛЫТЬ* IN GERMAN AND RUSSIAN

The article presents a contrastive analysis of motivational structures in German and Russian movement verbs, focusing on derivational forms of *schwimmen* and *плыть*. The research objective is to identify systemic correlations between prefixed verbs in the two languages, considering their semantics and motion orientation. This research perspective is determined by investigating morpho-semantic asymmetry and the degree of grammaticalization of spatial relations in German and Russian.

The methodology employs corpus analysis (DWDS, DGD, Ozhegov's dictionary), componential semantic decomposition and a contrastive approach. The German language exhibits 18 prefixed verbs derived from *schwimmen*, which have been classified into semantic

clusters. The comparative analysis revealed divergent encoding strategies for spatial semantics: German prefixes convey more precise spatial information, whereas Russian verbs predominantly require contextual disambiguation.

The study establishes that German demonstrates a higher degree of grammaticalization in spatial relations, while Russian compensates through prefixal syncretism and pragmatic adaptability. The findings are systematically presented in a comparative table that highlights lexical lacunae in cross-linguistic correspondences. Future research directions include a diachronic analysis of verb usage evolution. The present research both develops theoretical models of motion representation and addresses practical challenges in translation practice and language teaching methodologies.

Key words: motion verbs; prefixes; German National Corpus; lacunarity; linguistic topology; contrastive linguistics.

Глаголы движения представляют собой важный пласт лексики в любом языке, отражая фундаментальные способы концептуализации пространственного перемещения. В данной статье мы сосредотачиваемся на сопоставительном анализе мотивационных структур немецких и русских глаголов движения, в частности на формантах мотивирующей основы *schwimmen* и *плыть* в некоторых значениях физического перемещения. Глаголы движения вызывают научный интерес у ряда исследователей [2; 4; 5], с точки зрения их семантических признаков [6, с. 38; 7, с. 136]. **Целью** исследования становится выявление систематизированной связи между немецкими и русскими глаголами на основе их приставочной семантики и, как следствие, направленности их движения. **Актуальность** работы обусловлена необходимостью изучения морфо-семантической асимметрии немецкого и русского языков, а также степени грамматикализации пространственных отношений. **Материалом** исследования послужили глаголы, взятые из корпуса немецкого языка цифрового словаря DWDS [9] и толкового словаря русского языка С. И. Ожегова [3]. Методы исследования включают корпусный анализ, компонентное разложение семантики, а также контрастивный анализ.

Анализируя приставочные формы глагола *schwimmen*, выявленные методом сплошной выборки из словаря немецкого языка DWDS, обнаружено 18 глаголов со значением физического перемещения. Исследуя представленные глаголы с точки зрения семантики, мы считаем целесообразным сгруппировать их в несколько кластеров:

1. Прохождение расстояния по произвольной траектории:
 - а) движение к объекту (*anschwimmen, erschwimmen, herüberschwimmen, hinschwimmen, zurückschwimmen, zuschwimmen*);
 - б) удаление от объекта (*abschwimmen, davonschwimmen, fortschwimmen, hinausschwimmen, rausschwimmen, wegschwimmen*);
 - в) движение через/вокруг/мимо (*durchschwimmen¹, hinüberschwimmen, runterschwimmen, vorbeischwimmen, umherschwimmen*);
 - г) движение по вертикальной оси (*aufschwimmen*);
2. Глаголы, обозначающие стиль плавания (*brustschwimmen, delfinschwimmen, rückenschwimmen*);

3. Продолжение движения (*durchschwimmen*², *weilerschwimmen*);
4. Совместное движение / сопровождение (*mitschwimmen*).

В толковом словаре С. И. Ожегова [3] выявлено наличие 11 префиксальных глаголов, некоторые из которых полисемичны, что суммарно дает 15 семантических значений, которые также интегрируются по определенным признакам в несколько кластеров:

1. Прохождение расстояния по произвольной траектории:
 - а) движение к объекту (*доплыть*, *подплыть*, *приплыть*, *наплыть*¹);
 - б) удаление от объекта (*заплыть*¹, *отплыть*, *уплыть*);
 - в) движение через/вокруг/мимо (*переплыть*, *проплыть*¹, *оплыть*);
 - г) движение по вертикальной оси (*всплыть*);
2. Наличие преграды (*заплыть*², *наплыть*², *проплыть*², *выплыть*²).

Коррелируя немецкоязычные префиксальные глаголы с русскоязычными приставочными эквивалентами предлагается провести структурно-семантический анализ формантов. В рамках проводимого исследования фокус внимания направлен на глаголы из кластера '*Прохождение расстояния по произвольной траектории*' ввиду его идентичности в русском и немецком языках, а, как известно, иная лингвокультура всегда воспринимается через призму собственной лингвокультуры [1, с. 125]. Однако приставочные глаголы, составляющие подгруппы данного кластера отличаются в русском и немецком языках, что определяет направление сопоставительного анализа в рамках данного исследования.

Для прецизионной характеристики семантического значения немецкоязычных формантов мотивирующей основы *schwimmen* были использованы дефиниции из немецкоязычного словаря Duden, сопровождаемые примерами из корпуса немецкого языка цифрового словаря DWDS, а также из корпуса DGD Института немецкого языка имени Лейбница (Leibniz-Institut für Deutsche Sprache) в Мангейме. Рассмотрим кластер '*Прохождение расстояния по произвольной траектории*' с выделенными в нем подкластерами в зависимости от семантических значений:

а) движение к объекту

- **anschwimmen** – «auf etwas zuschwimmen, zu etwas hinschwimmen» '*подплыть к чему - либо / кому - либо*': Einen zischenden Schwan, der vom Holzsteg anschwamm und in seine Schuhe beißen wollte, vertrieb er mit Steinen [9] '*Шипящего лебедя, который подплыл к деревянному мостку и пытался укунить его за ботинки, он отогнал камнями*';

- **erschwimmen** – «beim Schwimmen erringen» '*завоевывать / достигать какой - либо точки, плавая*'; акцент на результате: Ihre Steigerung im Vergleich zur Deutschen Meisterschaft in Berlin hat aber auch die lange Freistilstaffel bitter nötig, um eine Medaille zu erschwimmen [9] '*Однако достигнутый прогресс по сравнению с результатами Чемпионата Германии в Берлине свидетельствует о том, что эстафетной команде вольного стиля также необходимо существенное улучшение показателей, чтобы доплыть до медали*';

- – «von der gegenüberliegenden Seite hierher [zur sprechenden Person] schwimmen» *‘доплыть до говорящего с противоположной стороны’*: Der Oberförster im polnischen Krosno hat sogar vermutet, dass die Tiere aus Deutschland über die Oder herübergeschwommen seien [9] *‘Главный лесничий города Кросно в Польше даже предположил, что животные приплыли Из Германии через Одер’*;

- **hinschwimmen** – «an eine bestimmte Stelle schwimmen» *‘доплыть до определенной точки’*; акцент на целенаправленном движении к какой-либо точке: ...da gab es so eine kleine Insel und wer da halt hinschwimmen wollte, musste einen von den Rettungsschwimmern mitnehmen [8] *‘...там был небольшой остров, и если кто-то хотел до него доплыть, требовалось сопровождение одного из спасателей’*;

- **zuschwimmen** – «sich schwimmend (auf jemanden, etwas) zubewegen» *‘направляться вплавь к чему-то / кому-то’*; акцент на целенаправленном движении к объекту: ...das ist ein irres Gefühl, wenn die Haie so auf dich zuschwimmen und du ihr Maul von unten sehen kannst [8] *‘...невероятное ощущение, когда акулы плывут прямо на тебя и снизу ты видишь их открытые пасти’*;

- **zurückschwimmen** – «wieder an den, in Richtung auf den Ausgangsort schwimmen» *‘возвращаться вплавь к исходной точке’*: ...du hast der untere Torpedo haut ja nur ab, du musst zurückschwimmen [8] *‘...нижняя торпеда просто отваливается – ты должен плыть обратно’*.

б) удаление от объекта

- **abschwimmen** – «sich schwimmend entfernen, wegschwimmen» *‘отплывать от объекта’*; акцент на отдалении от исходной точки: Die aufgestaunten Eisschollen sind in Richtung Oderhaff abgeschwommen, so das Wasserstraßen- und Schifffahrtsamt Eberswaide gestern [9] *‘Скопившиеся ледяные глыбы отплыли в направлении Одерского залива, как сообщило вчера Управление водных путей и судоходства Эберсвальде’*;

- **davonschwimmen** – «sich schwimmend entfernen» *‘удаляться от объекта вплавь’*; акцент на полном удалении от объекта: Doch ich war schon von Anfang an gaumensicher und wußte es wie die Schildkröte, die ausschlüpft und ins Meer davonschwimmt [9] *‘Но я изначально был неуязвим для критики и осознавал это, подобно черепашке, которая, едва вылупившись, уплывает в море’*;

- **fortschwimmen** – «sich schwimmend entfernen; wegschwimmen» *‘уплывать / отдаляться от объекта вплавь все дальше’*; акцент на продолжении движения в направлении от исходной точки, без указания конкретного пункта назначения): Auch fielen möglicherweise einige Insassen der Rettungsboote Haien zum Opfer, als sie von den Booten fortschwammen [9] *‘Не исключено, что часть людей из спасательных шлюпок была атакована акулами, когда они отплывали подальше от шлюпок’*;

- **hinausschwimmen** – «sich schwimmend vom Ufer entfernen» *‘отдаляться от берега, плывя’*: Er hatte die Hütte mitten in der Nacht verlassen und war ins Meer hinausgeschwommen... [9] *‘Он вышел из хижины поздней ночью и отплыл в море...’*;

- **rausschwimmen** – «hinausschwimmen» (разг.) *‘отдаляться от берега, плывя’*: Bei schönem Wetter ist es leicht rummelig, muß man ganz schön weit rausschwimmen, um ungestört schnorcheln zu können [9] *‘В хорошую погоду здесь многолюдно, поэтому приходится довольно далеко отплывать от берега, чтобы спокойно понырять с маской’*;

- **wegschwimmen** – «fortschwimmen» *‘уплыть, удаление от исходной точки / объекта’*; акцент на исчезновении из поля зрения, плывя: Ich blieb aber auch ganz cool, fühlte mich nur nicht so, und wartete, bis er wegschwamm, aber er schwamm nicht weg [9] *‘Внешне я оставался невозмутимым – хоть и не чувствовал себя таким – и ждал, пока он уплывет, но он не уплывал’*.

в) движение через/вокруг/мимо

- **durchschwimmen**¹ – «sich schwimmend durch, unter, zwischen etwas hindurchbewegen» *‘двигаться сквозь / под / между чем-либо, плывя’*: Immerhin kann man den See noch umrunden und muss nicht durchschwimmen [9] *‘По крайней мере, озеро можно обойти по берегу и проплыть его насквозь не обязательно’*;

- **umherschwimmen** – «sich ziellos im Wasser bewegen» *‘плыть в водном пространстве бесцельно, хаотично’*: Das Leben der Fische, die scheinbar zeitlos und emotionslos in dem kleinen Becken umherschwimmen, spielt sich ab vor einer Tapete, die wohl ein Korallenriff abbilden soll [9] *‘Рыбы, бесчувственные и существующие вне временного хронотопа, кружат хаотично в небольшом аквариуме перед обоями, тщетно пытающимися изобразить коралловый риф’*;

- **vorbeischwimmen** – «in die Richtung einer Person, Sache schwimmen und ohne sie zu berühren in der gleichen Richtung (wieder von ihr weg) weiterschwimmen» *‘проплывать мимо кого-то / чего-то, не соприкасаясь, и продолжать движение в том же направлении’*: Für uns waren die sogenannten «auswärtigen Angelegenheiten» wie Treibhölzer, die im Fluss vorbeischwimmen... [9] *‘Для нас эти так называемые «внешнеполитические дела» были как проплывающие мимо коряги...’*;

- **hinüberschwimmen** – «nach drüben schwimmen» *‘переплыть на другой берег’*: Er will zu ihr hinüberschwimmen; sie stellt ein Licht auf, um ihm im Wasser die Richtung zu zeigen [9] *‘Он хочет переплыть к ней на другой берег; она ставит светильник, чтобы указать ему путь в воде’*;

- **runterschwimmen** – «in einem fließenden Gewässer mit der Strömung schwimmen» *‘плыть вниз по течению’*: Junge Dame, wissen Sie, dass Sie es mit einem der besten Spieler zu tun hatten, der Mississippi runterschwamm? [9] *‘Молодая леди, знаете ли вы, что имели дело с одним из лучших игроков, что когда-либо плыли по Миссисипи вниз по течению?’*.

г) движение по вертикальной оси

- **aufschwimmen** – «infolge des Auftriebs an die Wasseroberfläche kommen»
‘всплывать на поверхность воды под действием подъемной силы’: Während des Tages hatten Bergungstrupps Gas in den Schiffskörper gepumpt, damit der Tanker aufschwimmen konnte [9] ‘Спасательные бригады в течение суток закачивали газ в отсеки судна, чтобы танкер смог всплыть’.

Семантический анализ представленных глаголов

Таким образом, проведенный семантический анализ позволяет провести сравнительный семантический анализ русскоязычных и немецкоязычных формантов от мотивирующих основ соответствующих глаголов *плыть* и *schwimmen* и объединить их в сводную таблицу:

Сопоставительная таблица подкластеров семантических значений
немецкоязычных и русскоязычных формантов мотивирующих основ
глаголов *плыть* и *schwimmen* кластера
‘Прохождение расстояния по произвольной траектории’

Направление	Немецкие глаголы	Русские глаголы
К объекту	anschwimmen	подплыть
	erschwimmen	доплыть, прилагая усилия
	hinschwimmen	доплыть до цели
	herüberschwimmen	приплыть с противоположной стороны
	zuschwimmen	—
	zurückschwimmen	—
	—	наплыть ¹
От объекта	abschwimmen	отплыть
	hinausschwimmen	отплыть от берега
	rausschwimmen (разг.)	отгребать от берега (разг.)
	davonschwimmen	уплыть от исходной точки
	fortschwimmen	уплывать дальше
	wegschwimmen	уплывать вовсе
	—	заплыть ¹
Через/вокруг/мимо	vorbeischwimmen	проплыть мимо
	durchschwimmen ¹	проплыть сквозь / между
	hinüberschwimmen	переплыть
	umherschwimmen	—
	runterschwimmen (разг.)	—
	—	оплыть
Вертикально	aufschwimmen	всплыть

Корреляция немецкоязычных и русскоязычных формантов демонстрирует отсутствие соответствий у некоторых глаголов. Так, у глаголов из подкластера «движение к объекту», а также из подкластера «движение через /вокруг /мимо» наблюдается отсутствие выраженных приставочными глаголами от основы *плыть* русскоязычных эквивалентов. Например, глаголы *zuschwimmen* переводится на русский язык как ‘плыть к’, *zurückschwimmen* ‘плыть обратно’, *umherschwimmen* ‘плыть хаотично’ и *runterschwimmen* ‘плыть вниз по течению’. В данном случае перевод на русский язык возможен исключительно непосредственно глаголом *плыть* с соответствующими предлогами, что свидетельствует о высокой степени лексикализации пространственных значений в немецком языке, т. е. немецкоязычные префиксы кодируют более точную информацию в формантах по сравнению с русским языком.

Напротив, в русском языке присутствуют форманты, *наплыть*¹ (движение к объекту), *заплыть*¹ (движение от объекта) и *оплыть* (движение через /вокруг /мимо), которые на немецкий язык переводятся глаголами, не содержащими мотивирующую основу *schwimmen*.

Важно отметить, что в русском языке некоторые форманты многозначны, например, *заплыть*¹ «переместиться вглубь водного пространства, удаляясь от исходной точки» [3, с. 215] и *заплыть*² «отклониться от исходной траектории движение по водному пространству и оказаться за каким-либо препятствием» [3, с. 215]. При этом русскоязычных формантов по сравнению с немецкоязычными наблюдается в полтора раза меньше.

Таким образом, немецкий язык демонстрирует бóльшую грамматикализацию пространственных отношений, в то время как русский язык, сохраняя синкретизм приставок, активизирует контекстуальные механизмы для передачи нюансов. Перспективой дальнейшего исследования становится изучение частотности употребления представленных глаголов в зависимости от различных временных периодов.

ЛИТЕРАТУРА

1. Ларина Т. С. Имплицитная директивность в поликодовом тексте // Когнитивные исследования языка. 2019. № 38. С. 124–130.
2. Майсак Т. А., Рахилина Е. В.. Глаголы движения и нахождения в воде : лексические системы и семантические параметры. М. : Индрик, 2007. 752 с.
3. Ожегов С. И., Шведова Н. Ю. Толковый словарь русского языка. М. : А ТЕМП, 2006. 938 с.
4. Сулейманова О. А., Ильясова Д. А. Лингвотопологическая схематизация движения глаголами идти и бежать в русско-английском пере воде // Вестник МГПУ. Серия «Филология. Теория языка. Языковое образование». 2025. № 1(57). 120–132.
5. Сулейманова О. А., Соколова А. В. Методологический потенциал фрейм-ового моделирования и языковой картины мира : дивергенция и конвергенция // Terra Linguistica. 2024. Т.15, № 4. С. 123–133.

6. Хлопова А. И., Ладоса О. М. Диагностика динамики значения базового концепта Sicherheit (безопасность) в немецкой лингвокультуре (на материале корпуса устных текстов и данных ассоциативного эксперимента) // Научный журнал «Современные лингвистические и методико-дидактические исследования». 2021. №3 (51). С. 35–46.
7. Хлопова А. И., Ладоса О. М. Лексико-семантическое содержание ценности "Erfolg" в немецкой лингвокультуре // Теоретическая и прикладная лингвистика. 2021. Т.7, № 2. С. 135–151.
8. DGD – Die Datenbank für Gesprochenes Deutsch [Electronic resource]. URL: <https://dgd.ids-mannheim.de/> (date of access: 01.07.2025).
9. DWDS – Digitale Wörterbuch der deutschen Sprache [Electronic resource]. URL: <https://www.dwds.de/> (date of access: 02.07.2025).

УДК 811.111'367+81'33+004.91

Nuriyeva Saadat Tapdıq

Doctor of Philosophy in Philological

Sciences, Associate Professor

Azerbaijan University of Languages

Baku, Azerbaijan

saadetnuriyeva@gmail.com

THE SYNTAX OF DOUBLE MODALITY IN STANDARD ENGLISH: A CORPUS-BASED STUDY

This article presents a corpus-based study of the syntactic structure and grammatical behavior of double modality in Standard English. While double modal constructions such as *might could* or *may can* are typically associated with non-standard dialects, Standard English employs syntactically permissible combinations involving a core modal and a semi-modal or lexical modal phrase (e.g., *might be able to*, *may be able*, *might have to*, *may have to*, *could be going to*, “*must be able to*”, etc.). The study draws on data from contemporary English corpora to identify and analyze these combinations, it also aims to analyze these combinations from a syntactic perspective, focusing on their linear order, internal structure, and the distributional rules that govern them. By combining insights from corpus analysis and modern grammatical theory, the article identifies productive and frequent patterns of double modality and distinguishes them from dialectal or regionally restricted forms. It also explores the interaction between modality, tense, and aspect in multi-modal constructions. The corpus evidence reveals that although Standard English avoids true double modals, it permits layered modal expressions through grammatically structured auxiliary chains, reflecting complex semantic layering. The findings contribute to a clearer understanding of modality stacking and provide corpus-driven insights into the syntactic flexibility and evolving usage patterns of modal expressions in contemporary English. Corpus Examples of Double Modality in Standard English (core modal + lexical/semi-modal such as “*might be able to*”, “*may have to*”, “*could be going to*”, “*must be able to*”, “*might want to*”) obtained from sources such as the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), and the NOW Corpus, and adapted to reflect real and standard sentence structures.

Key words: syntax; double modality; a corpus-based study; standard English; core modal; periphrastic modal.

In recent decades, modality has remained a central focus in linguistic research due to its crucial role in expressing speakers' attitudes, degrees of certainty, obligation, and possibility. While single modal auxiliaries such as *can*, *must*, or *may* have been extensively studied in both formal and functional approaches, less attention has been given to double modality—constructions that involve two modal elements within the same clause. Double modality refers to the use of two modal elements in a clause to express layered modal meanings—e.g., possibility + ability (*might can*), or obligation + necessity (*should have to*).

Traditionally associated with non-standard dialects of English, such as Southern American English or Scots, double modals like *might could* or *may can* have often been viewed as regional or informal phenomena.

However, Standard English also exhibits forms of double modality, albeit in more grammatically sanctioned structures. These typically involve a combination of a core modal (e.g., *might*, *may*) and a semi-modal or lexical modal (e.g., *be able to*, *have to*, *want to*), as seen in constructions such as *might be able to* or *may have to*. While these forms may not constitute true double modals in the dialectal sense, they represent a type of modal layering that reflects syntactic and semantic complexity.

This article investigates the syntactic structure and distributional behavior of such double modality patterns in Standard English. Drawing upon data from contemporary English corpora and informed by modern syntactic theory, the study aims to explore how these modal combinations are constructed, what syntactic constraints govern them, and how they differ from or relate to non-standard forms. Special attention is given to issues such as *linear order*, *auxiliary stacking*, and *the interaction between modality, tense, and aspect*.

By focusing on corpus-based evidence, this research contributes to a more comprehensive understanding of modality in English, particularly the phenomenon of modality stacking within the bounds of standard grammar. The findings not only expand the theoretical discussion around modals but also shed light on the syntactic flexibility and expressive richness of modal constructions in contemporary usage.

The methods applied in the article's analysis

1. Corpus-Based Analysis

The study draws on authentic language data from major English corpora such as the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), and the NOW Corpus.

These corpora are used to extract and examine naturally occurring examples of double modality in Standard English.

The corpus data provide evidence of real usage patterns, frequency, and syntactic environments of core modal + lexical/semi-modal combinations.

Syntactic Analysis

The article employs descriptive syntactic analysis **to** examine the internal structure, linear order, and auxiliary stacking within modal combinations.

It focuses on syntactic constraints, distributional rules, and the interaction between modality, tense, and aspect in multi-modal constructions.

Comparative Theoretical Approach

The study integrates contemporary grammatical theory to compare standard and non-standard forms of double modality.

It distinguishes productive, grammatically sanctioned patterns in Standard English from dialectal or regionally restricted ones.

Theoretical background

In non-standard varieties of English, particularly Southern American English and Scots English, so-called *true double modals* are commonly used. These constructions involve the sequential use of two modal verbs within a single clause, allowing speakers to express nuanced combinations of meanings such as possibility, ability, permission, or necessity.

Leading researchers on double modals in English dialects include Michael Montgomery, Margaret Mishoe, Marianna Di Paolo, Barbara Fennell, Stephen Nagle and et al.

One of the most prominent scholars who has investigated this phenomenon is **Michael Montgomery**. According to Montgomery (1989), although combinations such as *might could*, *may can*, or *should oughta* are considered ungrammatical in Standard English, they function as established and meaningful structures in certain regional dialects. He traces the origins of these double modals to 18th-century Scots English, suggesting that the forms were preserved in Southern U.S. English through historical migration patterns, particularly by Scots-Irish settlers.

Battistella (1995) investigates the syntax of the so-called double modal constructions as they appear in Southern United States English. The paper analyzes why and how these structures—such as *might could* or *may can*—exist in non-standard dialects, situating them within a broader syntactic variation framework.

Battistella (4-int. resources) proposes a model where generative grammars consist of a core and a periphery, whereby the “extra” (spurious) modal is licensed similarly to adverbials. This approach extends existing theories of how adverbs are licensed and supports C. L. Baker’s view that negative ordering involves peripheral grammatical mechanisms, applying it to double modals.

The study not only clarifies the syntactic properties and idiosyncratic behavior of double modals but also considers their implications for current proposals regarding adverbials and negation. By treating these modal combinations as structurally licensed extensions of peripheral syntax, Battistella’s analysis offers new insights into regional syntactic variation. (1995)

Anita Fetzer’s study (2010, 2012, 2014, 2018) focused on Pragmatics and modality in English. While her primary focus is on pragmatics, Fetzer has analyzed modality from a discourse perspective, exploring how modal expressions contribute to meaning in context.

Collins, Peter also investigated modals and Quasi-modals in English (2009). His contribution is that he studied both core modals and semi-**modals**, emphasizing their syntactic and distributional properties in various registers of English, including formal and informal varieties.

However, unlike non-standard English, in **standard English** this linguistic phenomenon (double modality), although it exists as we mentioned earlier, has not been thoroughly studied.

Nuriyeva, S. T. has approached double modality from a distinct perspective. For the first time, she has analyzed double modality specifically in Standard English. Her published articles address the phenomenon from both structural and semantic angles, employing componential analysis to explain the internal structure and meaning of double modal constructions.

Nureyeva, S. T. (2022) presented inversion as a means of expressing the category of modality in her article “*Müasir ingilis dilində inversiya modalıq kateqoriyasının ifadə vasitəsi kimi*” (“Inversion as a Means of Expressing the

Category of Modality in Modern English"). By accepting inversion as a means of expressing modality, she argues that inversion, along with other modal devices, contributes to the formation of double modality at the sentence level of the language. **Nuriyeva, S. T. (2024)** also presents "The Particle as an independent part of speech, based on the classifications of well-known scholars, and considers it a component of double modality. **Nuriyeva, S. T. (2010, 2023, 2024)** analyzes double modality from componential, structural, and semantic perspectives. According to Nuriyeva, S. T., in Standard English, double modality is formed by a simple modal verb combined with modal verb equivalents—namely, periphrastic modals (multi-word constructions ending in *to*). These are expressed through the following structural patterns.

Table 1

Ways of expression of double modality in Standard English
may+be allowed to
may+have to
might+have to
might+be able to might+want to
must+be able to
should+have to
should +be able to
could+be going to

Yet modern standard English is rich in examples related to this topic. For this reason, our research is based on examples culled from dictionaries. The following are corpus examples of double modality used in standard English.

They should be able to find you the best package to suit your needs.

CollinsCobuild (En-En)

You may have to learn a few new skills along the way. CollinsCobuild (En-En)

You may have to work part-time, in a badly paid job with unsociable hours...

Science may be able to provide some explanations of paranormal phenomena. CollinsCobuild (En-En)

Just hang on a tick, we may be able to help... CollinsCobuild (En-En)

There's a good chance that we may be able to help. OxfordDictionary (En-En)

She could be going to start her own business soon.

You might want to check your email again.

Periphrastic modals function semantically like true modals.

The following table shows the modal verbs along with their corresponding periphrastic modals. The examples following the table support this idea.

Table 2

Modal verbs	Periphrastic modals
Can	to be able to
Will	to be going to, to be about to
Must	to have to, to have got to
Should	ought to, to be to, to be supposed to
May	to be allowed to, to be permitted to

You may be allowed to claim a refund of the tax. (OxfordDictionary (En-En))

He may have to pay a fine. CollinsCobuild (En-En)

Just hang on a tick, we may be able to help... CollinsCobuild (En-En)

There's a good chance that we may be able to help. OxfordDictionary (En-En)

She may be able to squeeze you in, if you play your cards right.
OxfordDictionary (En-En)

They may be able to help with childcare so that you can have a break...
CollinsCobuild (En-En)

He conjectured that some individuals may be able to detect major calamities...
CollinsCobuild (En-En)

We may be able to design a course to suit your particular needs...
CollinsCobuild (En-En)

If you need to return to the UK quickly, British Consular officials may be able
to arrange it. CollinsCobuild (En-En)

You might want to check your email again. CollinsCobuild (En-En)

The syntax of double modality (Standard English)

After identifying the existence of double modality in standard English and its
means of expression, we should now focus on its syntax.

What is meant by “Syntax of double modality”?

The syntax of double modality refers to the structural organization and
grammatical constraints governing the combination of two modal elements in a
single clause, including their linear order, compatibility, and contribution to the
clause's tense, aspect, and modality interpretation.

1. What is meant by Leaner Order?

It means the order of core modal and semi modal. In standard English core
modal always precedes the semi- modal (periphrastic modal). There are examples
and analysis by Leaner Order:

1. He might be able to help.

Leaner order: might (core modal) → *be able to* (semi-modal)

Meaning: **might** expresses possibility, **be able to** expresses “ability”.

So: He might be able to help. = It is possible that he has the ability to help.

2. She may have to leave early.

Leaner order: may (core modal) → *have to* (semi-modal)

Meaning: **may** expresses possibility, **have to** expresses “Necessity”.

So: She may have to leave early. = It is possible that she will be required to leave early.

3. She must be going to announce it soon.

Must (core modal) → *be going to* (future intention marker)

Linear order: *must* (core modal) → *be going to* (future intention marker)

Meaning: *must* expresses certainty + *be going to* expresses intention

So: She must be going to announce it soon.=It is certain they are planning to announce it soon.

The preferred linear order in Standard English for expressing double modality is:

Core modal verb + a semi-modal or lexical modal

You could want to reconsider your decision.

Linear order: *could* (core modal) → *want to* (lexical modal expression)

Meaning: Possibility + Desire (“It is possible that you might desire to reconsider.”)

The following also refers to the syntax of double modality:

2. The Sentence Structure in which Double Modals Appear

One of the main features of the syntax of double modality is its position within the sentence structure. Since word order in English is fixed, the predicate (the verb phrase) always follows the subject. When the predicate is expressed by a double modal construction, it occupies the second position in the sentence. What distinguishes this type of predicate from others is that it consists of three components: a core modal, a semi-modal (or periphrastic modal), and an infinitive. This structure expresses double meaning. E.g.,

We might need to help. Linear order: *might* (core modal) → *need to* (lexical/modal equivalent verb) Meaning: Possibility + Obligation *We might need to help* = *It is possible that we are required to help.*

3. The grammatical roles of modals.

"Grammatical roles" refers to their function in the sentence—such as expressing modality (ability, necessity, permission, etc.), acting as auxiliary verbs, and affecting the tense or aspect of the verb phrase. In standard English the first component of the double modal changes according to the category of tense. For example;

He may have to go there.

He might have to go there.

It goes without saying that the two examples differ according to their meaning.

4. The syntactic function of the double modals in the sentence.

Double modals, together with the infinitive, serve the syntactic function of what we term a **double modal predicate**. We believe this label best captures the structure in question. This is due to the fact that the first element is a **core modal**,

the second is a **semi-modal** or a **modal expression**, and the third is the **infinitive**. Together, these three components fulfill the syntactic role of the **predicate** in relation to the **subject** of the sentence.

5. Double Modals in the negative

It should be noted that in standard English, only the first modal (the core modal) is used in the negative. The second element—the periphrastic modal or modal equivalent—is **not** negated.

Example:

She might not have to leave early.

might – core modal (possibility)

not – negation

have to – periphrastic modal (obligation)

leave – main verb (infinitive)

Conclusion

This paper has explored the syntactic and semantic features of double modality in Standard English, with a focus on constructions involving a core modal followed by a periphrastic or lexical modal expression. While traditionally considered a phenomenon limited to non-standard dialects, our analysis demonstrates that Standard English employs its own grammatically sanctioned forms of modal layering, such as *might have to*, *may be able to*, and *should want to*.

Using data from major English corpora, we have identified consistent structural patterns in these constructions. The findings confirm that the core modal always precedes the semi-modal, forming a linear and rule-governed predicate that expresses layered meanings—typically combining modality types such as possibility, necessity, ability, or intention.

The syntactic analysis shows that these double modality structures function as complex predicates and obey standard English grammatical constraints, including fixed word order and auxiliary behavior. Additionally, we observed that negation applies exclusively to the core modal in such constructions, further supporting their structural regularity.

By integrating corpus-based evidence with syntactic theory, this study contributes to the growing body of research on modal expressions in English. It challenges the assumption that double modality is solely a dialectal feature and highlights the expressive potential and syntactic flexibility of Standard English. These findings open pathways for further research on modality stacking and its role in encoding subtle distinctions in speaker stance and communicative intention.

REFERENCES

1. Battistella, E. L. (1995). *The syntax of the double modal construction*. *Linguistica Atlantica*, 17, 19–44.

2. Battistella, E. Negation in double modal constructions. 1991. *Linguistic Analysis* 21:49–65.
3. Boertien, et al. Constituent structure of double modals. 294. 1986. University of Alabama
4. Di Paolo, Marianna. Double modals as single lexical items. 1989. 64:195–224.
5. Fetzer, A. (2014). *Evidentiality in discourse. Intercultural Pragmatics*, 11(3), ... DOI:10.1515/ip-2014-0015.
6. Fetzer, A. (2018). *Communicative action meets discourse analysis: Discourse pragmatics*. In C. Ilie & N. R. Norrick (Eds.), *Pragmatics and its Interfaces* (pp. 33–57). John Benjamins.
7. Fetzer, A., & Johansson, M. (2010/2012). *Cognitive verbs in context*. In *Corpus Studies in Contrastive Linguistics* (pp. 89–115). John Benjamins.
8. Mishoe, M., & Montgomery, M. (1994). *The pragmatics of multiple modal variation in North and South Carolina. American Speech*, 69(1), 3–29.
9. Montgomery, M. B. (1989). *Exploring the roots of Appalachian English. English World-Wide*, 10(2), 227–278.
10. Montgomery, M. B. (1989). *Exploring the roots of Southern modal constructions. English World-Wide*, 10(2), 129–142.
11. Nagle, Stephen, 1994 “The English double modal conspiracy” *Diachronica* 11:199–212.
12. Nuriyeva S. T. “QOŞA MODALLIĞIN TƏDQIQI TARIXINDƏN” “Yeni dövrdə təhsil və tədqiqat fəaliyyəti: reallıqlar və çağırışlar” II Beynəlxalq elmi konfransı, 13–14 dekabr 2024-cü il, Mingəçevir Dövlət Universiteti, Azərbaycan. Səh. 333–336.
13. Nuriyeva S. T. MÜASİR İNGİLİS DİLİNDƏ QOŞA MODALLIĞIN YARANMASINDA PERIFRASTİK MODAL FƏİLLƏRİN ROLU. Quba respublika elmi konfransı 20 dek. 2024. Abbasquluağa Bakıxanovun 230 illik yubileyinə həsr edilmiş konfrans. pp. 161–165.
14. Nuriyeva S. T. NEW APPROACHES TO SCIENTIFIC THEORETICAL IDEAS ABOUT THE PART OF SPEECH CALLED "PARTICLE" IN MODERN ENGLISH. Norwegian Journal of development of the International Science №147/2024. pp. 64–67.
15. Nuriyeva S. T. “Müasir ingilis dilində inversiya modallıq kateqoriyasının ifadə vasitəsi kimi” filologiya məsələləri, №14, 2022. səh. 191–201.
16. Nuriyeva S. T. “Müasir ingilis dilində inversiya modallıq kateqoriyasının ifadə vasitəsi kimi” filologiya məsələləri, №14, 2022. səh. 191–201.
17. Nuriyeva S. T. İngilis və Azərbaycan dillərində modallıq kateqoriyasının qarşılıqlı-müqayisəli təhlilinə dair Azərbaycan Kəpublikası Elm və Təhsil Nazirliyi. Beynəlxalq Elmi Konfrans (13–14 aprel 2023 cü il. Sumqayıt Dövlət Universiteti. Konfrans materialları. 2023 №1. s. 389–392.
18. Nuriyeva S. T. Nitqdə qoşa modallığın komponental təhlili. Ümummilli Lider Heydər Əliyevin anadan olmasının 100-cü ildönümünə həsr olunmuş “Ali təhsildə dayanıqlı inkişaf məqsədləri” mövzusunda Beynəlxalq elmi-praktik konfransı. 11 aprel 2023-cü il, s. 299–300.

19. Nuriyeva S. T. Qoşa modallığın komponental təhlili Publisher Agency, The 2nd International Scientific Conference Progress in Science» (April 13–14, 2023). Brussels, Belgium.
20. Nuriyeva S. T. “Qoşa modallıq” ADU, Elmi xəbərlər, Bakı-2010, № 5, səh. 102–112.
21. Nuriyeva S. T. Müasir ingilis dilində qoşa modallıq gələcək zamanın ifadə vasitəsi kimi. Bakı, AMEA, Tədqiqlər I, 2019.

DICTIONARIES

1. OxfordDictionary (En-En)
2. CollinsCobuild (En-En)

INTERNET RESOURCES

1. <https://www.english-corpora.org/coca/> (COCA – Contemporary American English)
2. <https://corpus.byu.edu/bnc/> (British National Corpus)
3. <https://skell.sketchengine.eu/>
4. multimodal.library.sc.educambridge.org+2journals.lib.unb.ca+2readpaper.com+2.

Секция 2. ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РЕШЕНИЯ ЛИНГВИСТИЧЕСКИХ ЗАДАЧ

УДК 81'32, 81'33

Астапенко Елизавета Валентиновна
студент
Санкт-Петербургский
государственный университет
г. Могилев, Беларусь

Lizaveta Astapenka
Student
St. Petersburg State University
Mogilev, Belarus
li.astapenka@gmail.com

Митрофанова Ольга Александровна
кандидат филологических наук,
доцент
Санкт-Петербургский
государственный университет
г. Санкт-Петербург, Россия

Olga Mitrofanova
PhD in Philology,
Associate Professor
St. Petersburg State University
Saint Petersburg, Russia
o.mitrofanova@spbu.ru

ЛЕКСИКО-СЕМАНТИЧЕСКИЕ ОТНОШЕНИЯ В КОРПУСАХ И ЯЗЫКОВЫХ МОДЕЛЯХ ДЛЯ БЕЛОРУССКОГО ЯЗЫКА

В данной работе исследуется способность языковых моделей фиксировать семантические связи между словами. Описаны эксперименты по автоматическому предсказанию лексико-семантических отношений в *белорусском* языке с помощью различных моделей, а именно Word2Vec, BERT и генеративных языковых моделей. Качество предсказаний моделей определялось с помощью процедуры псевдодизамбигуации (Word2Vec) и экспертной оценки (BERT, Gemini 2.5 Pro). Полученные результаты могут быть применены для решения различных задач: для создания и обогащения лексических баз данных, исследования сочетаемости слов, улучшения машинного перевода, перефразирования, реферирования и других систем автоматической обработки белорусского языка.

Ключевые слова: лексико-семантические отношения; белорусский язык; Word2Vec; BERT; большие языковые модели.

LEXICAL-SEMANTIC RELATIONS IN CORPORA AND LANGUAGE MODELS FOR THE BELARUSIAN LANGUAGE

Our study examines the ability of language models to capture semantic relations between words. We describe experiments on automatic prediction of lexical-semantic relations in the *Belarusian* language using various models (Word2Vec, BERT, and generative language models). The performance of the models was evaluated by pseudo-disambiguation test (Word2Vec) as well as by expert evaluation method (BERT, Gemini 2.5 Pro). The results can be applied to create and enrich lexical databases, to analyse word co-occurrence, to improve machine translation, paraphrasing, summarization, and other systems related to automatic processing of the Belarusian language.

Key words: lexical-semantic relations; the Belarusian language; Word2Vec; BERT; large language models.

Современные нейросетевые модели позволяют решать разнообразные задачи в области обработки естественного языка и исследовать не только морфологические и синтаксические, но и семантические свойства языковых выражений. Распространение моделей распределенных векторных вложений расширяет возможности автоматического распознавания лексико-семантических отношений в корпусах текстов, что особенно актуально для языков, для которых в открытом доступе еще не представлены лексические базы данных и компьютерные тезаурусы. Для белорусского языка существуют различные онлайн-словари (например, Verbum [1], Slounik [2]), инструменты для морфологической и синтаксической разметки (Stanza [3], UD-Pipe [4]). В 2013 г. в рамках проекта «Экспериментальный корпус белорусского языка» был реализован лемматизатор YABC_Tagger [5]. На платформе Corpus.by [6] представлены разработанные в лаборатории распознавания и синтеза речи ОИПИ НАН Беларуси сервисы для автоматической обработки текстов на белорусском языке. В институте языкознания имени Якуба Коласа проводятся работы по созданию Национального корпуса («Беларускі N-корпус») [7], Грамматической базы белорусского языка и т. д. Самые известные NLP-инструменты для белорусского языка представлены в репозитории [8]. Однако на сегодняшний день в открытом доступе не размещены языковые модели, обученные на белорусских корпусах и предназначенные для решения задач лексической семантики. Таким образом, актуальность нашего исследования определяется необходимостью восполнения лакун в области развития общедоступных автоматизированных систем, позволяющих выявлять семантические отношения в лексике, а также большим диапазоном задач в компьютерной лингвистике, для решения которых такие модели могут быть использованы. Цель нашей работы состоит в обучении и оценке языковых моделей белорусского языка для предсказания регулярных парадигматических и синтагматических отношений, наблюдаемых в корпусах текстов.

Для обучения языковых моделей был сформирован корпус белорусского языка. Текстовая коллекция включает в себя три сегмента:

- корпус белорусского языка в проекте Universal Dependencies (UD) объемом 25 231 предложение (305 417 токенов), содержащий художественные, публицистические тексты, тексты Википедии и социальных медиа [9];
- Belacorus объемом 246 текстов (1 535 047 токенов), включающий тексты различных жанров за период 1987–2010 гг. [10];
- корпус белорусского языка из коллекции Лейпцигского университета (Web 300K) объемом 300 000 предложений, источником которого служат различные веб-сайты на белорусском языке [11].

Итоговый объем корпуса составил 413 012 предложений (5 662 829 токенов). Объединенный корпус представлен в двух версиях: корпус без разметки (для возможности отбора контекстов) и размеченный корпус (для дальнейшего обучения моделей). Корпус UD имел предварительную морфосинтаксическую разметку, лемматизация и частеречная разметка корпусов Belacorus [10] и Web-300K-2015 [11] была проведена нами с помощью библиотеки Stanza [3].

Для проведения экспериментов по дифференциации синонимов и антонимов был сформирован набор данных, состоящий из 1939 синонимических рядов, которые были извлечены из онлайн-словаря синонимов [12], а также 597 антонимических пар из словаря «Слоўнік лексічных формаў (сінонімы, амонімы, антонімы, паронімы, амографы, амафоны)» [13]. Для повышения сбалансированности данных набор антонимов был расширен за счет автоматического подбора к ним синонимов, благодаря чему общее количество антонимических пар увеличилось до 1624.

На первом этапе предсказание лексико-семантических отношений проводилось с помощью моделей Word2Vec [14]. В результате серии экспериментов был произведен выбор оптимальных параметров обучения моделей skip-gram (vector_size = 250, window = 5, min_count = 5, epochs = 5) и cbow (vector_size = 300, window = 5, min_count = 5, epochs = 5). Оценка качества предсказаний опирается на процедуру псевдодизамбигуации [15], по результатам которой модели достигают высоких значений точности для списков как с целевым существительным, так и с прилагательным (78 % и 90 % для модели skip-gram, 83 % и 92 % для модели cbow). Для визуализации работы моделей на платформе Hugging Face с помощью библиотеки Streamlit было разработано приложение [16]. Пример предсказаний моделей приведен в таблице.

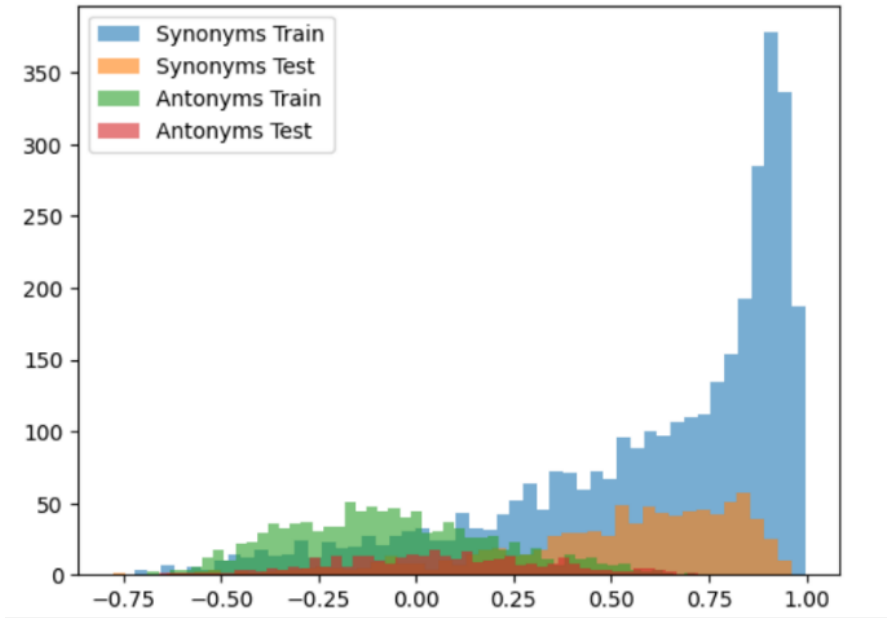
Пример выходных данных для слова *мова* (язык)

skip-gram, ADJ	Cos	CBOW, NOUN	Cos
<i>англійські_ADJ</i>	0,593	<i>праванис_NOUN</i>	0,589
<i>руські_ADJ</i>	0,592	<i>філалогія_NOUN</i>	0,584
<i>українські_ADJ</i>	0,578	<i>літаратура_NOUN</i>	0,562
<i>царкоўнаславянскі_ADJ</i>	0,574	<i>арфаграфія_NOUN</i>	0,552
<i>старабеларускі_ADJ</i>	0,554	<i>маўленне_NOUN</i>	0,54

Сиамские нейронные сети

Для улучшения качества предсказаний использовался подход, объединяющий обращение к дополнительным лексическим ресурсам и применение сиамских нейронных сетей. За основу был взят алгоритм Siamese Semantic Vectors [17] с компонентом Word2Vec skip-gram, более чувствительным к контексту по сравнению с cbow. В качестве обучающих данных выступали списки синонимов и антонимов для белорусского языка, составленные по материалам словарей. Общее число пар синонимов составило 4038, антонимов – 1624. Эти данные были разделены на обучающую и тестовую выборки в процентном соотношении 80:20.

Для оценки моделей была использована метрика, определяющая долю от всех синонимов (антонимов), которую модель смогла рассчитать правильно. Были введены две границы так, что верными считаются все предсказания косинусной меры от 0,6 до 1 для синонимов и от -0,2 до -1 для антонимов. Результаты представлены на рисунке.



Распределение значений косинусного сходства
для синонимических и антонимических пар после обучения нейросети

Приведем в качестве примера изменение значения косинусного сходства для пары синонимов *палахлівы_ADJ* (пугливый) – *баязлівы_ADJ* (трусливый), а также для пары антонимов *смелы_ADJ* (смелый) – *баязлівы_ADJ* (трусливый). После обучения нейронной сети оно увеличилось с 0,8713 до 0,9293 в первом случае и уменьшилось с 0,5167 до -0,4429 во втором.

Дообучение модели BERT

Способность моделей BERT к построению контекстуализированных эмбедингов оказывается особенно полезной при автоматическом подборе синонимических замен слов в предложениях. Задачу предсказания моделью синонимов можно свести к задаче маскированного языкового моделирования, при котором BERT предлагает наиболее вероятные слова вместо специального токена [MASK] в последовательности. Следовательно, необходимо научить модель предсказывать для конкретного контекста не просто возможные варианты заполнения позиции маскированного токена, а именно синонимичные. В этом эксперименте тренировочные данные включили в себя 1203 синонимические пары. Кроме того, были составлены тестовые данные объемом в 100 предложений (по 25 контекстов на 4 основные части речи (NOUN, ADJ, VERB, ADV)).

Для белорусского языка существует несколько многоязычных и одноязычных предобученных моделей, среди которых для дальнейших экспери-

ментов была выбрана модель *roberta-small-belarusian* (1 язык, 15.7 млн параметров) [18], показавшая лучшие результаты в тестах. При дообучении модели гиперпараметры принимали следующие значения: *num_train_epochs* = 6 (число эпох), *per_device_train_batch_size* = 16 (размер батча), *learning_rate* = $5e-5$ (скорость обучения).

Дополнительные методы улучшения качества предсказаний предполагают сравнение эмбедингов предложений с помощью Sentence Transformer (SBERT) с предварительным добавлением словарных синонимов к списку уже подобранных моделью кандидатов. Поскольку для белорусского языка модели вида SBERT отсутствуют, были выбраны многоязычные модели LaBSE (110 языков, 471 млн параметров) [19] и *paraphrase-multilingual-MiniLM-L12-v2* (50 языков, 118 млн параметров) [20]. На тестовой выборке было доказано преимущество первой модели.

Итоговый алгоритм включает в себя нескольких этапов:

- на вход дообученная модель *roberta-small-belarusian* получает предложение со спецтокеном [MASK] и предсказывает возможные варианты, которые могли бы стоять на его месте;
- к списку также добавляются кандидаты из словаря, которые были найдены в нем по лемме целевого слова, полученной с помощью библиотеки Stanza;
- на место маски в предложение по очереди подставляются варианты замен из списка;
- модель LaBSE сравнивает эмбединги предложений и ранжирует кандидатов в зависимости от значения косинусной меры.

В качестве примера выходных данных можно привести предложенные моделью синонимические замены для целевого слова *асаблівасць* в контексте *Была ў Фёдара яшчэ адна рэдкая [MASK]: на сваёй уласнай ініцыятыве ён ніколі нікога не падвозіў, не падбіраў*:

- *асаблівасць*: 1,0000
- *адметнасць*: 0,9986
- *характарыстыка*: 0,9974
- *уласцівасць*: 0,9972
- *своеасаблівасць*: 0,9961.

Для оценки качества предсказаний модели до процедуры дообучения и после нами были проведены два опроса. В анкеты были внесены тестовые предложения вместе с полученными для них ответами модели. Все данные были разделены случайным образом на пять выборок. Для оценки того, насколько модель улавливает семантические связи между словами, экспертам нужно было поставить баллы от 0 до 5. В качестве экспертов выступали носители белорусского языка. Общее количество информантов для каждой из анкет составило 25 человек. На основании полученных результатов определялась согласованность экспертов, средняя и относительная оценки. Значение согласованности варьируется от 0,75 до 0,88, что интерпретируется как

значительный уровень согласованности мнений экспертов [21]. По результатам оценки, проведенной для тестовых выборок в целом и всех предложений в отдельности, можно сделать вывод о том, что задача предсказания синонимов в контексте для белорусского языка с помощью модели BERT решается с достаточно высокой точностью.

Генерация синонимов с помощью больших языковых моделей

В связи с распространением больших языковых моделей (LLM) мы решили оценить их способность предсказывать синонимы в контексте для белорусского языка. Работа моделей проверялась на составленном нами тестовом наборе данных, включающем в себя 100 предложений. Среди многоязычных больших языковых моделей были выбраны наиболее известные и доступные в настоящее время, а именно DeepSeek-V3, GPT-4o, Claude 3.5 Haiku, Gemini 2.0 Flash, Gemini 2.5 Pro. Затем был составлен базовый промпт на белорусском языке и получены ответы моделей. На основе анализа ошибок и совпадений предложенных вариантов со словарными данными для дальнейшей экспертной оценки использовалась модель Gemini 2.5 Pro [22]. Тестовые данные были разделены на 5 выборок по 20 предложений. Экспертам предлагалось поставить 0 или 1 для каждого варианта в зависимости от того, может ли он быть использован в качестве замены в предложении. Для всех выборок рассчитывалась средняя оценка, а также согласованность экспертов (с помощью коэффициента каппа Флейса, как и в случае с моделями BERT). В группу экспертов вошли носители белорусского языка (84 человека). Согласованность принимает значения в диапазоне от 0,49 до 0,59, что является достоверным результатом для гуманитарных областей [21]. Средняя оценка варьируется между 50 и 60 баллами для всех выборок, что свидетельствует о том, что модель не в полной мере справляется с задачей генерации синонимов в контекстах для белорусского языка.

В результате исследования было проведено обучение языковых моделей разных архитектур для белорусского языка и их оценка в задаче предсказания парадигматических и синтагматических отношений, наблюдаемых в корпусах текстов. Прделанная работа позволяет заложить основу для последующей разработки более сложных лингвистических систем, охватывающих разнообразные семантические связи в белорусском языке. Обученные и протестированные нами модели могут быть не только интегрированы в подобные системы, но и применены в таких областях, как информационный поиск, машинный перевод, лингводидактика и т. д.

ЛИТЕРАТУРА

1. Verbum [Electronic Resource]. URL: <https://verbum.by/> (дата обращения: 15.07.2025).
2. Slounik. URL: <https://slounik.org/> (дата обращения: 15.07.2025).
3. Stanza. URL: <https://stanfordnlp.github.io/stanza/index.html> (дата обращения: 15.07.2025).

4. UD-Pipe. URL: <https://ufal.mff.cuni.cz/udpipe/2/models> (дата обращения: 15.07.2025).
5. YABC_Tagger. URL: https://github.com/poritski/YABC_Tagger/blob/master/docs/readme_BE.md (дата обращения: 15.07.2025).
6. Corpus.by. URL: <https://corpus.by/index.php?lang=be> (дата обращения: 15.07.2025).
7. Беларускі N-корпус. URL: <https://bnkorporus.info/> (дата обращения: 15.07.2025).
8. be_nlp_speech_resources. URL: https://github.com/navalnica/be_nlp_speech_resources (дата обращения: 15.07.2025).
9. Shishkina Y., Lyashevskaya O. Sculpting enhanced dependencies for Belarusian // Analysis of Images, Social Networks and Texts: 10th International Conference. AIST 2021. Cham: Springer, 2022. Vol 13217. P. 137–147.
10. Belacorporus. URL: <https://github.com/Belarusian-Corporus> (дата обращения: 15.07.2025).
11. Belarusian Corpora. Wortschatz.uni-leipzig.de. URL: <https://wortschatz.uni-leipzig.de/en/download/Belarusian> (дата обращения: 15.07.2025).
12. Клышка М. К. Слоўнік сінонімаў і блізказначных слоў / пад рэд. Л. А. Антанюк. 2-е выд., выпр. і дапоўн. Менск: Вышэйшая школа, 1993. 445 с.
13. Хвале́й Я. І., Ша́рпіла У. В. Слоўнік лексічных формаў (сінонімы, амонімы, антонімы, паронімы, амографы, амафоны). Мн.: Парадокс, 2004. 416 с.
14. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. 2013. URL: <https://arxiv.org/abs/1301.3781> (дата обращения: 15.07.2025).
15. Petrushenko L., Mitrofanova O. Predicting Style-Dependent Collocations in Russian Text Corpora // Proceedings of Recent Advances in Slavonic Natural Language Processing (RASLAN 2023). 2023. P. 79–89.
16. Lexical-semantic Calculator for the Belarusian Language. <https://huggingface.co/spaces/lizaastapenka/lexical-semantic-calculator-for-the-Belarusian-language> (дата обращения: 15.07.2025).
17. Siamese Semantic Vectors. URL: <https://github.com/maxwelljohn/word-vector-remapping> (дата обращения: 15.07.2025).
18. roberta-small-belarusian. URL: <https://huggingface.co/KoichiYasuoka/roberta-small-belarusian> (дата обращения: 15.07.2025).
19. LaBSE. URL: <https://huggingface.co/sentence-transformers/LaBSE> (дата обращения: 15.07.2025).
20. Paraphrase-multilingual-MiniLM-L12-v2. URL: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2> (дата обращения: 15.07.2025).
21. McHugh M. L. Interrater reliability: the kappa statistic // Biochemia medica. 2012. Vol. 22. № 3. P. 276–282.
22. Gemini. URL: <https://deepmind.google/technologies/gemini/> (дата обращения: 15.07.2025).

УДК 004.8 + 81'367.3

Безменова Елизавета Сергеевна
преподаватель кафедры истории
и грамматики английского языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Elizaveta Bezmenova
Lecturer of History of English
and English Grammar Department
Belarusian State University
of Foreign Languages
Minsk, Belarus
bzmnva@gmail.com

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ДЛЯ ИНТЕРПРЕТАЦИИ КАНОНИЧЕСКИХ МЕНАСИВНЫХ ВЫСКАЗЫВАНИЙ

В статье исследуются возможности автоматической интерпретации речевых актов угрозы современными NLP-системами. На основе интегративного подхода, объединяющего методы лингвопрагматики и семантического синтаксиса с возможностями инструментария искусственного интеллекта, проводится анализ канонического менасивного высказывания бинарной модели, где комиссивный (намерение причинить вред) и директивный (условие выполнения угрозы) компоненты структурированы в условную конструкцию *if-then*. Экспериментальная часть исследования включает обработку типичного менасивного высказывания из телесериала «Игра престолов» тремя типами языковых моделей DeepSeekChat, ChatGPT и Telegram-чат-ботом. Разработанные промпты позволяют оценить способности систем к эмоционально-прагматической интерпретации угрозы, металингвистической рефлексии и генерации ситуативно-адекватных реакций. Результаты демонстрируют частичную успешность идентификации прямой угрозы при сохраняющихся трудностях распознавания контекстуально-обусловленных менасивов и воспроизведения социально-иерархических моделей поведения.

Ключевые слова: речевой акт угрозы; менасивные высказывания; искусственный интеллект; бинарная модель; каноническая конструкция; реактивная стратегия.

ARTIFICIAL INTELLIGENCE FOR THE INTERPRETATION OF CANONICAL THREATENING STATEMENTS

The article examines the potential for automatic interpretation of speech acts of threats within contemporary Natural Language Processing (NLP) systems. Employing an integrative approach that synthesises methodologies from linguopragmatics and semantic syntax alongside the capabilities of artificial intelligence tools, the analysis focuses on the canonical utterance with the meaning of threat as outlined in the binary model, where the commissive component, which pertains to the intention to cause harm, and the directive component, which represents the condition for the fulfillment of the threat, are structured within a conditional *if-then* construction. The empirical component of this study entails the examination of a prototypical utterance with the meaning of threat derived from the television series *Game of Thrones*, subjected to analysis by three distinct language models: DeepSeek Chat, ChatGPT, and a Telegram chatbot. The prompts developed for this research facilitate an assessment of the systems' capacities for emotional-pragmatic interpretation of threats, metalinguistic reflection, and the generation of contextually appropriate responses. The findings indicate a degree of success in the identification of direct threats; however, challenges persist in the recognition of contextually determined threats and the replication of social-hierarchical behavior models.

Key words: speech act of threat; utterances with the meaning of threat; artificial intelligence; binary model; canonical construction; reactive strategy.

Угроза как коммуникативный феномен продолжает оставаться одним из наиболее значимых и сложных элементов человеческого взаимодействия, поскольку пронизывает все уровни социальной жизни: от межличностных конфликтов до политических переговоров. В современную эпоху цифровой коммуникации, когда границы между прямыми и опосредованными формами общения становятся все более размытыми, менасивные паттерны приобретают особую значимость, выступая не только инструментом принуждения, но и важным маркером социальных отношений, власти и статуса. Отсутствие паралингвистических сигналов (мимики, жестов) и анонимность коммуникации создают новые барьеры для интерпретации угроз, делая их анализ актуальным как для лингвистики, так и непосредственно для сферы искусственного интеллекта. Вариативность восприятия угроз в зависимости от культурного контекста, статуса коммуникантов и канала общения **подчеркивает важность** разграничения канонических и девиантных форм менасивов, а также ставит вопрос о возможности их точной автоматической обработки.

Указанные особенности определяют междисциплинарный статус исследования речевых актов угрозы (далее – РАУ), объединяющего три научных области: лингвопрагматику, изучающую языковые явления во взаимосвязи с экстралингвистическими факторами (такими как статус коммуникантов и ситуативный контекст), семантический синтаксис, который обеспечивает инструментарий для анализа структур передачи менасивной семантики, и искусственный интеллект.

В работе используется интегративный подход для сравнительного анализа моделей восприятия и реактивных стратегий в NLP-системах при обработке РАУ. Такой ракурс позволяет не только оценить степень соответствия машинной интерпретации человеческим когнитивным стратегиям, но и выявить определенные системные закономерности обработки эксплицитных интенций в автоматизированных системах.

Исследование РАУ основывается на фундаментальных положениях семантико-синтаксического анализа, для которого приоритетной задачей является установление корреляции между номинативными элементами семантической структуры и их вербальным воплощением в высказываниях с менасивной семантикой. Мы исходим из базовой модели РАУ – бинарной структуры, подразумевающей наличие комиссивного компонента (выражение намерения причинить вред) и директивного компонента (условие выполнения угрозы), – основанной на лексикографическом анализе глагола *'threaten'* (Кембриджский, Оксфордский словари) [1, 2], предполагает обязательное наличие субъекта-агенса (адресанта угрозы) и объекта-пациенса (адресата).

В смоделированном на базе описанной конструкции высказывании со значением угрозы *I will cause trouble / hurt you if you do not do what I want* 'Я

причину тебе вред, если ты не сделаешь то, что я хочу' подлежащее главного предложения (*I*) выполняет семантическую роль агенса и соответствует адресанту угрозы, тогда как дополнение (*you*) выступает как пациент, репрезентируя адресата. Глагол *threaten* функционирует как ядерный элемент РАУ, организующий вокруг себя обязательные семантические роли, а каноническая структура представляет собой условную конструкцию *if-then* с четким разделением на комиссивную и директивную составляющие (главную и придаточную клаузы соответственно) [3].

Представленное высказывание демонстрирует идеальную реализацию модели, где синтаксические позиции, их семантические роли и прагматические функции находятся в строгом соответствии друг с другом, что служит методологической основой для исследования обработки менасивных высказываний NLP-моделями.

Методика исследования предполагает два последовательных этапа. Первый этап (лингвопрагматический): отбор тестового высказывания РАУ осуществляется на основе его соответствия канонической бинарной структуре, что подтверждается семантико-синтаксическим анализом распределения ролей и наличием ядерных глаголов угрозы. Лингвопрагматическое обоснование выбора учитывает статусные отношения коммуникантов, ситуативный контекст и потенциальный перлокутивный эффект. Второй этап (экспериментальный) посвящен тестированию NLP-систем посредством специально разработанных промптов, включающих два аспекта анализа: оценку эмоционально-прагматической нагрузки высказывания, а также моделирование поведенческих реакций, направленное на генерацию стратегий ответного речевого поведения в заданном коммуникативном контексте. Основным является анализ не столько сравнительной эффективности моделей, сколько их способности воспроизводить адекватные стратегии реагирования, сопоставимые с поведенческими паттернами человека. Особое внимание уделяется также выявлению характерных ограничений интерпретационных возможностей систем при обработке РАУ.

Первый этап. **В качестве менасивного высказывания для тестового эксперимента отобран пример**, взятый из телесериала «Игра престолов» (2011–2019), экранизации цикла Дж. Р. Р. Мартина: *And I swear this: if you ever betray me, I'll burn you alive* 'И клянусь: если ты когда-нибудь предашь меня, я сожгу тебя заживо'. Данная реплика представляет собой классическую реализацию менасивного речевого акта, где четко прослеживаются все структурные и прагматические характеристики угрозы. Так, семантико-синтаксический анализ позволяет выявить характерную бинарную организацию высказывания. Условная клауза *if you ever betray me* 'если ты когда-нибудь предашь меня' выполняет директивную функцию, устанавливая запрет на определенное действие (предательство), в то время как главная клауза *I'll burn you alive* 'я сожгу тебя заживо' реализует комиссивный компонент, в свою очередь эксплицируя конкретную санкцию за нарушение этого запрета. В данном высказывании сохраняется классическое распре-

ление семантических ролей, характерное для канонических менасивных конструкций. В главной клаузе адресант угрозы (Дейенерис) однозначно занимает позицию агенса как субъект, инициирующий насильственное действие, тогда как адресат (Варис) выступает в роли пациенса – объекта воздействия. В придаточной клаузе происходит семантическая инверсия: адресат становится агенсом потенциального действия (предательства), демонстрируя сознательный выбор; адресант переходит в позицию пациенса, временно становясь объектом этого действия. Такое динамичное перераспределение ролей акцентирует причинно-следственную связь (за предательством неминуемо следует казнь) и усиливает перлокутивный эффект угрозы, поскольку главная клауза закрепляет доминирование адресанта. Глагол *burn* ‘сожгу’ выступает ядерным элементом угрозы, а уточнение *alive* ‘заживо’ демонстрирует жестокость предполагаемого наказания, исключая любое метафорическое прочтение. Отметим и то, что перформативный глагол *swear* ‘клянусь’ придает высказыванию характер клятвы, повышая его иллокутивную силу и подчеркивая неизбежность реализации угрозы (подтвержденную последующей казнью Вариса).

Анализ прагматических характеристик данной коммуникативной ситуации позволяет утверждать, что высказывание построено по стратегии прямого иллокутивного воздействия. Использование экспрессивной лексики и отсылка к символическому для Таргариенов способу казни (сожжение) служат мощными маркерами, подчеркивающими серьезность намерений адресанта угрозы. Коммуникативные роли участников четко выражены: Дейенерис как доминирующий субъект власти и Варис как подчиненный адресат. Паралингвистические и кинетические особенности реализации высказывания демонстрируют контролируемую агрессию со стороны адресанта. Безэмоциональный тон королевы, ее ровная интонация и отсутствие повышения громкости сигнализируют о холодной расчетливости, что типично для ритуализированных угроз в иерархических отношениях. Медленное приближение к адресату создает эффект давления. Наблюдается характерная реакция Вариса, выражающаяся в последовательной смене эмоциональных паттернов: сначала хмурость, затем вынужденная улыбка. В сочетании с отсутствием движений и статичным положением рук это может быть интерпретировано как стратегическое подчинение. В данной ситуации невербальные маркеры дополняют вербальную угрозу, делая акцент на асимметрии власти через контроль пространства и подавление поведенческих реакций адресата. Как отмечается в описании сцены, Дейенерис не просто предупреждает, а именно обещает казнь через *Dracarys*, что также соответствует типичной для Таргариенов риторике абсолютной власти, где угрозы являются инструментом поддержания лояльности [4].

На основании проведенного лингвопрагматического анализа, рассматриваемого менасивного высказывания перейдем к описанию экспериментальной процедуры оценки его интерпретации несколькими NLP-моделями.

Второй этап. Для исследования возможностей искусственного интеллекта в интерпретации РАУ созданы два специализированных промпта, направленных на выявление поведенческих реактивных стратегий системы в ответ на менасивное высказывание и на анализ когнитивной интерпретации интенциональной природы речевого акта.

Первый промпт сформулирован как имитация реальной коммуникативной ситуации: «Привет! У меня к тебе вопрос: как бы ты отреагировал и что бы сделал, если бы тебе сказали следующее? And I swear this: if you ever betray me, I'll burn you alive (И клянусь: если ты когда-нибудь предашь меня, я сожгу тебя заживо)». Такая формулировка позволяет оценить, насколько модель способна генерировать социально приемлемые и ситуативно адекватные реакции на менасив.

Второй промпт ориентирован непосредственно на лингвистический анализ высказывания: «Каково твое восприятие данного высказывания? Считаешь ли ты это прямой угрозой или, скорее, косвенным речевым актом?». Данная формулировка направлена в свою очередь на выявление способности модели к металингвистической рефлексии и классификации РАУ по параметру прямой/косвенной коммуникации.

Настоящий эксперимент состоял из серии операций по обработке анализируемого высказывания тремя типами NLP-систем: генеративной моделью DeepSeekChat, диалоговой системой ChatGPT на архитектуре GPT-4 и узкоспециализированным Telegram-чат-ботом, что позволило получить сравнительные данные об особенностях интерпретации РАУ различными по своей архитектуре и функционалу искусственными интеллектуальными системами. **Для всех моделей использовались идентичные промпты с двумя аспектами анализа – оценкой эмоционально-прагматической нагрузки высказывания и моделированием поведенческих реакций. В ходе эксперимента проведено по 2 тестовые итерации для каждой системы, что позволило выявить устойчивые схемы шаблонных реакций, характерные для автоматизированного анализа РАУ.**

Результаты обработки анализируемого высказывания тремя NLP-моделями демонстрируют как общие закономерности обработки стратегий интерпретации угрозы, так и их вариативность, обусловленную архитектурными различиями систем.

DeepSeekChat продемонстрировал комплексный подход, сочетающий лингвистический анализ с прагматической оценкой. Модель идентифицировала высказывание как потенциально прямую угрозу (акцентируя перформатив *I swear* и конкретику насилия – *burn you alive*), но допустила возможность косвенной интерпретации при наличии контекстуальных маркеров (гипербола, художественный дискурс). См. рис. 1. Реактивная стратегия включала развернутый сценарий взаимодействия: от запроса уточнений («Что подразумевается под предательством?») до рекомендаций по обеспечению безопасности. Подобный подход демонстрирует попытку воспроизведения характерной для человека когнитивной гибкости.

3. Моё субъективное восприятие

Без контекста фраза звучит как **потенциально прямая угроза**, потому что:

- Используется клятва («/ swear»), что придаёт серьёзность.
- Конкретный способ расправы («сожгу заживо») — это не абстрактная злость, а образ пытки/убийства.

Но если это, например, цитата из фильма, игра или эмоциональный всплеск — тогда это **гиперболизированный косвенный акт**.

Рис. 1. DeepSeekChat

ChatGPT, несмотря на автоматическое удаление исходного запроса системой модерации (с пометкой о возможном нарушении политики использования), предоставил развернутый анализ, классифицировав высказывание как эмоционально-насыщенный речевой акт двойственной природы. Модель выделила оппозицию «буквальная угроза vs. метафорическое выражение», подчеркнув зависимость интерпретации от паралингвистических факторов (тон, невербалика). См. рис. 2. Предложенные моделью действия (установление границ, оценка эмоционального состояния говорящего) ориентированы на деэскалацию предполагаемого конфликта, что в целом соответствует социально-адаптивным схемам, типичным для диалоговых систем.

1. Прямая угроза или косвенный речевой акт?

- **Прямая угроза:**
Если воспринимать слова буквально, здесь есть чёткое обещание причинить серьёзный физический вред в ответ на предполагаемое действие (предательство). По форме это соответствует прямой угрозе — явному предупреждению о насилии.
- **Косвенный речевой акт:**
Однако часто в реальной жизни такие выражения могут быть гиперболизированы, эмоционально окрашены и служить для выражения глубины чувства обиды, страха или желания показать крайнюю степень серьёзности. Тогда это скорее **метафорическое выражение**, которое передаёт эмоциональный настрой, а не реальный план действий.

Рис. 2. ChatGPT

Telegram-чат-бот продемонстрировал упрощенную, но более категоричную позицию: высказывание однозначно трактовано как прямая угроза с акцентом на ее «жестокость», что отражает некоторую ограниченность узкоспециализированных ботов в обработке прагматических нюансов. Система предложила стандартный алгоритм действий (оценка контекста — установление границ — обеспечение безопасности), демонстрирующий определенную степень соответствия рекомендациям других NLP-моделей по параметру полноты охвата.

Примечательно, что, несмотря на различия в глубине анализа, все три модели предлагают схожий базовый сценарий реагирования, варьирующийся лишь степенью детализации и адаптивности (см. рис. 3).

4. Что делать с такой фразой?

— Если это угроза:

- Зафиксировать (например, скриншот в переписке).
- Чётко сказать, что такое неприемлемо: «*Ты сейчас серьёзно? Такие слова заставляют меня опасаться за свою безопасность*».
- Ограничить контакт.



DeepSeek Chat



ChatGPT 4.5 | DeepSeek | Midjourney

Привет! Это высказывание звучит очень угрожающе и может вызвать множество реакций. Если бы я был человеком, я мог бы отреагировать следующим образом:

1. **Оценка ситуации:** Прежде всего, я бы попытался понять контекст этого заявления. Почему человек так говорит? Есть ли у него основания для таких опасений?
2. **Обсуждение:** Я бы постарался поговорить с этим человеком, чтобы выяснить, что именно его беспокоит. Возможно, это просто эмоциональная реакция на что-то.
3. **Установка границ:** Если угроза воспринимается всерьез, я бы четко обозначил, что такие высказывания неприемлемы и не способствуют здоровым отношениям.
4. **Безопасность:** Если угроза кажется реальной, важно подумать о своей безопасности и, возможно, обратиться за помощью к друзьям или специалистам.
5. **Решение конфликта:** В зависимости от ситуации, может быть целесообразно рассмотреть возможность прекращения общения с человеком, если его поведение угрожает.

Telegram-чат-бот

Рис. 3. Сценарий реагирования интеллектуальными системами на каноническое менасивное высказывание

Сопоставление с оригинальной сценой из «Игры престолов», где адресат (Варис) демонстрирует стратегическое подчинение через невербальные маркеры (застывшая мимика, скованная поза с опущенными плечами и сцепленными перед собой руками) и вербальную уступчивость (*I would expect no thing less from the Mother of Dragons* ‘Я и не ожидал меньшего от Матери Драконов’), подчеркивает существенное расхождение с реакциями NLP-моделей. Ни одна из протестированных систем не воспроизвела иерархически-обусловленную модель поведения, характерную для ритуализированных угроз при асимметричных отношениях. Данный пробел обусловлен двумя факторами. Во-первых, это ограничение тренировочных данных — существующие корпуса для обучения NLP-систем в большинстве своем не содержат примеров вербального взаимодействия с четко выраженной иерархией участников. Во-вторых, модели не учитывают внеязыковые параметры

(социальные роли, историю взаимодействий, культурные коды), существенные для интерпретации РАУ, если данные аспекты не заданы явным образом в промпте.

Проведенное исследование позволило систематизировать центральные аспекты автоматической интерпретации канонических менаживных высказываний, включая способность современных NLP-систем к идентификации угроз (в нашем случае – канонической формы), их дифференциации на прямые/косвенные, а также генерации базовых социально адаптивных реакций, ориентированных на деэскалацию конфликта и обеспечение безопасности. Выявленные ограничения носят системный характер: шаблонность реакций, обусловленная недостаточностью тренировочных данных по сложным социальным взаимодействиям; упрощенная обработка прагматических нюансов, в том числе в контекстах иерархических отношений; проблемы модерации, приводящие к фильтрации исследуемого материала. Особенно заметны эти ограничения в узкоспециализированных системах, демонстрирующих негибкие стратегии реагирования.

ЛИТЕРАТУРА

1. Cambridge Dictionary [Electronic resource]. – Cambridge, 1999. – URL: <https://dictionary.cambridge.org/ru> (date of access: 12.07.2025).
2. Oxford Dictionary [Electronic source]. – London, 2000. – URL: <https://www.oxfordlearnersdictionaries.com> (date of access: 12.07.2025).
3. Карчевски Р. И. Прагмакоммуникативные характеристики языковых средств выражения угрозы в английском языке : дис. ... канд. филол. наук : 10.02.04. Минск, 2005. 122 с.
4. Abi-Khalil T. Game Of Thrones: 10 Quotes That Foreshadow Daenerys' Burning Of Kings Landing. Screen Rant [Electronic resource]. URL: <https://screenrant.com/game-of-thrones-10-quotes-that-foreshadow-daenerys-burning-of-kings-landing> (date of access: 13.07.2025).

УДК 81'322.2

Василевская Валерия Михайловна

аспирант кафедры теоретической
и прикладной лингвистики
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Valeria Vasilevskaya

PhD Student at the Department
of Theoretical and Applied Linguistics
Belarusian State University
of Foreign Languages
Minsk, Belarus
valeria-vas90@mail.ru

ОЦЕНКА КАЧЕСТВА РУССКОЯЗЫЧНЫХ АВТОМАТИЧЕСКИХ РЕФЕРАТОВ С ПРИМЕНЕНИЕМ МЕТРИК НА БАЗЕ N-ГРАММ

(на материале научных текстов из области ИКТ)

В данной статье проведен анализ применимости основанных на n-граммах метрик BLEU и ROUGE для оценки автоматически сгенерированных рефератов русскоязычных научных статей в области информационно-коммуникационных технологий (ИКТ). Выделены связанные со спецификой функционирования метрик проблемы, касающиеся русского языка и русскоязычных научных текстов, в частности, текстов из области ИКТ. Сделан вывод о недостаточной достоверности предоставленной данными метриками оценки качества русскоязычных научных рефератов ИКТ-текстов, связанной с лексико-морфологической сложностью русского языка и самой спецификой функционирования автоматических метрик на базе n-грамм.

Ключевые слова: автоматическое реферирование; информационно-коммуникационные технологии; научный дискурс; n-граммы; автоматические метрики; BLEU; ROUGE.

QUALITY EVALUATION OF RUSSIAN -LANGUAGE AUTOMATIC SUMMARIES WITH N-GRAM METRICS

(based on ICT Scientific Texts)

This article analyzes the applicability of n-gram-based metrics (BLEU and ROUGE) for evaluating automatically generated summaries of Russian-language scientific papers in the field of Information and Communication Technologies (ICT). The problems associated with the specifics of the metrics' functioning, concerning the Russian language and Russian-language scientific texts, in particular, texts from the field of ICT, are identified. The study concludes that the quality assessment of Russian-language scientific summaries of ICT texts, as provided by these metrics, lacks sufficient reliability. This limitation stems from the lexico-morphological complexity of the Russian language and the inherent constraints of n-gram-based automatic evaluation metrics.

Key words: automatic summarization; information and communication technologies; scientific discourse; n-grams; automatic metrics; BLEU; ROUGE.

Экспоненциальный рост объема научных публикаций обуславливает популярность автоматического реферирования как инструмента обработки научной информации. Особую актуальность оно приобретает в быстроразвивающихся дисциплинарных областях, к числу которых относятся информационно-коммуникационные технологии (ИКТ), где важен оперативный анализ

значительных массивов текстовой информации. В академической среде автоматическое реферирование способно ускорить обработку больших массивов публикаций, помогая исследователям быстрее находить релевантные работы.

В контексте активного внедрения нейросетевых технологий для автоматического реферирования в научно-исследовательскую практику особую актуальность приобретает проблема оценки и контроля качества генерируемых рефератов. Так, эталонный реферат научного текста должен:

1) сохранять содержательную адекватность: точная передача ключевых положений исходного текста при сохранении смысловой целостности и научной достоверности информации обеспечивает корректное восприятие научного контента;

2) соблюдать нормативы академического стиля: сохранение точной терминологии, логической последовательности изложения и соответствие жанровым нормам научной коммуникации.

Рефераты, генерируемые большими языковыми моделями, являются преимущественно абстрактными. Такой тип реферирования слабо поддается контролю со стороны пользователя. В этой связи особую актуальность приобретает задача оценки качества автоматических абстрактных рефератов на русском языке, характеризующемся высокой степенью синтетичности морфологии и синтаксиса, а также гибким порядком слов в предложениях.

Для автоматического контроля качества автоматического реферирования исследователями активно применяются метрики, основанные на *n*-граммах (последовательностях, состоящих из *n* элементов, которые могут быть звуками, слогами, словами или буквами, в зависимости от контекста), такие как BLEU и ROUGE. Спецификой данного типа метрик является: 1) учет локальных последовательностей (*n*-граммы анализируют последовательности из *n* подряд идущих элементов (слов, символов), что позволяет оценивать устойчивые статистические зависимости или повторяющиеся структуры, которые проявляются на ограниченных участках текста, обычно в пределах небольшого числа последовательно идущих элементов (слов, букв)); 2) статистическая природа получаемых результатов.

Эмпирическую базу исследования составили два реферата отобранной из электронной научной библиотеки «КиберЛенинка» статьи, посвященной методам автоматического анализа темпоральных выражений в текстах на естественном языке: 1) созданный вручную эталонный реферат, наличие которого является непременным условием для работы данных метрик и проверки их валидности (далее – эталон); 2) автоматический реферат, сгенерированный большой языковой моделью DeepSeek-V3 (далее – реферат).

Выбор статьи осуществлялся по следующим критериям: 1) соответствие тематической области ИКТ; 2) большой объем (45 341 знак); 3) наличие элементов, типичных для научных текстов в ИКТ-дискурсе: формул, таблиц, алгоритмов, насыщенность аббревиатурами преимущественно на латинице.

Инструментарий исследования представлен автоматическими метриками BLEU и ROUGE.

Целью данного исследования является определение эффективности и ограничений применения основанных на n-граммах метрик для оценки качества автоматически генерируемых рефератов русскоязычных научных текстов в области ИКТ.

Принципиальным **условием** исследования явилось моделирование типичной пользовательской ситуации при автоматическом реферировании: отсутствие предварительной обработки исходного текста, минимально детализированный запрос к модели (без составления сложных промптов), отсутствие постобработки сгенерированного реферата.

В ходе исследования сгенерированный нейросетью реферат был оценен автоматически по метрикам BLEU и ROUGE. Далее полученные по каждой из метрик данные подверглись анализу в контексте того факта, что изначально эти метрики были разработаны для английского языка, имеющего структуру принципиально отличную от русского языка. Необходимость использования данных метрик обусловлена отсутствием лингвистически адаптированных для русского языка метрик оценки качества автоматических рефератов.

1. Метрика BLEU (Bilingual Evaluation Understudy)

Данная автоматическая метрика была изначально разработана специалистами IBM в 2002 г. для оценки качества машинного перевода [1], но сегодня часто находит свое применение и в решении других задач генерации текста, в частности, оценке качества автоматического реферирования. Она сравнивает сгенерированный текст с эталонными референсами (в случае оценки качества автоматического реферирования эталоном будет созданный вручную реферат) на основе n-граммной статистики и оценивает точность совпадения n-грамм между автоматическим рефератом и эталоном. Эта метрика работает с последовательностями слов длиной 1–4 (униграммы (отдельные слова), биграммы (два последовательно идущих слова), триграммы (три последовательно идущих слова), кватрограммы (четыре последовательно идущих слова), чувствительна к пропускам и перестановкам слов, неточным формулировкам, а также штрафует за слишком короткие формулировки.

Данные расчетов по метрике BLEU составили (табл. 1):

Т а б л и ц а 1

Оценка автоматического реферата по метрике BLEU

Метрика	Оценка
Униграммы	0,75
Биграммы	0,62
Триграммы	0,55
Кватрограммы	0,40
ИТОГО	0,58

Метрика BLEU ранжируется от 0 при полном несовпадении до 1 при абсолютной идентичности с эталоном. Необходимо отметить, что с учетом строгости метрики при оценке абстрактивного реферата показатель 1 по умолчанию недостижим. Исходя из этого, результат 0,58 уже свидетельствует о формально приемлемом качестве сгенерированного реферата.

Важным будет рассмотреть примеры участков текста, получивших по данной метрике высокий, средний и низкий балл.

Пример 1: Высокий балл (0,7–1,0)
(эталон) *Стандарт ISO-TimeML утвержден ISO в 2009 году*
(реферат) *ISO-TimeML стал международным стандартом в 2009 г.*

Результат:

Униграммы: все слова присутствуют

Биграммы: 3 совпадения (*ISO-TimeML, 2009 г., международным стандартом*)

Триграммы: 1 совпадение (*в 2009 году ≈ в 2009 г.*)

Кватрограммы: 0 совпадений (*стандарт ISO-TimeML утвержден ≠ ISO-TimeML стал международным*)

Штраф: небольшой штраф за сокращение *году → г.*

Итоговый BLEU: ~0,65.

Интерпретация результата: несмотря на перефразирование, ключевые элементы сохранены.

На уровне униграмм, биграмм и триграмм работа BLEU вполне корректна. Далее при переходе к анализу кватрограмм проявляется особенность метрики, связанная с ее высокой чувствительностью к перефразированию: фактически предложение из автоматического реферата полностью корректно, однако высокий балл оно получает за исключительно техническое совпадение некоторого количества элементов без учета их положения в предложении и смысла, который они несут.

Пример 2: Средний балл (0,3–0,6)
(эталон) *Система HeidelTime достигает F-меры 86 % на Task A TempEval-2*
(реферат) *Точность HeidelTime составляет 86 %*

Результат:

Униграммы: 2 совпадения (*HeidelTime, 86 %*)

Биграммы: 1 совпадение (*86 %*)

Триграммы: 0 совпадений

Кватрограммы: 0 совпадений

Пропущено: *F-мера, TaskATempEval-2*

Итоговый BLEU: ~0,35.

Интерпретация результата: наблюдается потеря ключевых терминов.

В данном случае применение строгой метрики позволило обеспечить контроль над сохранностью терминологического состава: метрика абсолютно верно зафиксировала некорректную замену *F-меры* на *точность* (*F-мера* представляет собой среднеравновесное между точностью и полнотой). Однако следует подчеркнуть, что это актуально преимущественно в отношении англоязычных заимствований и аббревиатур, обладающих фиксирован-

ной формой. В отличие от них, русскоязычные термины подчиняются морфологическим законам русского языка, предполагающим их грамматическое изменение в соответствии с синтаксическим и семантическим контекстом конкретного высказывания.

Пример 3: Низкий балл (0–0,2)

(эталон) *Для русскоязычных текстов отсутствуют размеченные корпуса, аналогичные TimeBank.*

(реферат) *Для русского языка нет обучающих данных.*

Результат:

Униграммы: 1 совпадение (*Для*)

Итоговый BLEU: 0,1.

Интерпретация результата: практически полное несоответствие.

Здесь низкий балл обусловлен совпадением только одной униграммы. Возникает явный диссонанс: формально низкий балл за точность при фактической корректности предложения в сгенерированном реферате, хоть и содержащего обобщение (*размеченные корпуса, аналогичные TimeBank → обучающих данных*).

2. Группа метрик ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

Данная группа метрик была предложена ученым в области информатики Chin-Yew Lin [2] в 2004 г. с целью приблизить качество автоматической оценки рефератов к качеству ручного оценивания. В отличие от BLEU, ориентированной на точность, ROUGE оценивает полноту (recall) охвата ключевой информации из исходного текста на основе подсчета совпадений n-грамм (последовательностей слов) между рефератом и эталоном.

В настоящем исследовании из группы метрик были использованы ROUGE-1 (совпадение униграмм (отдельных слов)), ROUGE-2 (совпадение биграмм (пар слов)), ROUGE-L (совпадение длинных подпоследовательностей слов – непрерывных или разрывных цепочек слов, сохраняющих порядок следования и встречающихся как в автоматическом реферате, так и в эталоне).

Оценочные показатели ROUGE для автоматического реферата представлены следующим образом (табл. 2):

Таблица 2

Оценка автоматического реферата
по автоматическим метрикам ROUGE

Метрика	Содержание метрики	Оценка
ROUGE-1	совпадение униграмм	0,78
ROUGE-2	совпадение биграмм	0,65
ROUGE-L	совпадение длинных подпоследовательностей слов	0,72

Как и для BLEU, показатели данной группы метрик также стремятся к 1, по умолчанию недостижимой при оценке качества абстрактного реферата. Согласно группе метрик ROUGE автоматический реферат, имеющий достаточно высокие показатели, является качественным.

Однако при детальном анализе принципа работы метрик обнаруживаются существенные нюансы.

ROUGE-1 технически подсчитывает совпадение униграмм вне зависимости от окружающего их контекста.

Пример 1:

(эталон) *дейктические элементы*

(реферат) *обработка дейктических элементов*

Результат: униграмма *дейктические* засчитана.

Интерпретация результата: высокий показатель сходства (0,78) в данном случае свидетельствует исключительно о частичном совпадении униграмм сгенерированного реферата с эталонным без учета семантической и синтаксической корректности расположения единиц. Очевидно, что система распознает основы слов.

С учетом высокой флективности русского языка, существует значительная вероятность ошибочной положительной идентификации морфологически некорректных словоформ в автоматически сгенерированном реферате, что может привести к искусственному завышению оценки качества текста.

Тексты научного дискурса отличаются повышенной синтаксической и морфологической сложностью, терминология в области ИКТ характеризуется смешением русского и английского языков и высокой степенью аббревиации, что создает дополнительные трудности для метрических систем и повышает вероятность некорректной оценки качества автоматического реферата.

ROUGE-2 технически анализирует биграммы вне контекста.

Пример 2:

(эталон) *интервальная логика Аллена*

(реферат) *логика временных интервалов*

Результат: совпадающих биграмм нет.

Интерпретация результата: 0 баллов, однако фактически данные выражения являются контекстуальными синонимами, сохраняющими идентичный смысл.

Оценка (0,65) показывает наличие в тексте автоматического реферата некоторого количества совпадающих пар слов. При этом такой показатель не гарантирует корректность их употребления в реферате и значимость для понимания смысла исходного текста.

Свойственная русскому языку лексико-синтаксическая вариативность и отсутствие фиксированного порядка слов при изначальной англоязычной ориентированности метрики детерминируют высокую вероятность некор-

ректной идентификации значимой информации, а также учета биграмм, имеющих в составе служебные слова. Это, в свою очередь, может обусловить неверную оценку качества автоматического реферата.

ROUGE-1 + ROUGE-2. В некоторых случаях в работе двух сопряженных метрик наблюдается противоречие при оценке одного и того же участка текста.

Пример 3 (уровень предложения):

(эталон) Система *HeidelTime* использует правило-ориентированный подход.

(реферат) *HeidelTime* применяет правила.

Результат: ROUGE-1 даст высокий балл за *HeidelTime* и правила, но ROUGE-2 снизит балл за пропуск биграммы *правило-ориентированный подход*.

ROUGE-L технически оценивает длину наибольшей общей подпоследовательности. Метрика не требует строгого непрерывного совпадения. Это обеспечивает большую устойчивость к небольшим перефразированиям или добавлению служебных слов. ROUGE-L учитывает сохранение последовательности ключевых элементов, что важно для научных текстов, где логика изложения критична.

Пример 4:

(эталонная последовательность) Описание проблем. – Методы. – Примеры систем

(последовательность в реферате) Проблемы. – Методы. – (Примеры опущены)

Результат: частичное совпадение структуры → балл снижен.

Интерпретация результата: результат 0,72 свидетельствует о том, что технически структура автоматического реферата приблизительно соответствует структуре эталона.

Анализ принципа работы автоматических метрик ROUGE и BLEU позволяет выделить ряд проблем, связанных со спецификой функционирования метрик, основанных на статистическом учете n-грамм.

Проблемы, обусловленные спецификой русского языка

1. Гибкий порядок слов в русском языке снижает точность n-граммных метрик, так как они жестко привязаны к последовательности элементов.

2. Морфологическая вариативность детерминирует занижение оценки BLEU из-за несовпадения словоформ, даже если лексическая база идентична.

3. Контекстуальная синонимия и многозначность терминов не учитываются ни BLEU, ни ROUGE, что приводит к необоснованному занижению оценок при корректном перефразировании.

Проблемы, обусловленные применением метрик для научных текстов

4. В научном тексте допускается вариативность формулировок, но метрики, основанные на n-граммах, снижают оценки за отклонения от эталонного реферата.

5. Оценка исключительно лексического покрытия без учета смысловой связности нецелесообразна для научного реферата, требующего соблюдения

четкой структуры и логики. Также статистические метрики не различают стилевую окраску слов, что критично для оценки качества реферата научной статьи.

Проблемы, обусловленные спецификой текстов в области ИКТ

6. Научные тексты в области ИКТ характеризуются высокой терминологической плотностью, многие составные термины имеют сложную структуру. Термины такого типа BLEU может посчитать ошибочными при перефразировании или изменении морфологической формы.

7. Технические формулы и код: Ни BLEU, ни ROUGE не приспособлены для оценки вставок математических выражений или фрагментов программного кода, которые часто встречаются в ИКТ-рефератах.

Анализ выявленных проблем позволяет сделать **вывод** о том, что метрики, разработанные для английского языка с его фиксированным порядком слов и меньшей морфологической вариативностью, не предоставляют достоверную информацию о качестве русскоязычного научного реферата, включая даже базовые статистические показатели: оценки могут быть завышены либо занижены. Это связано с лексико-морфологической сложностью русского языка и самой спецификой функционирования автоматических метрик на базе n-грамм.

Для адекватной оценки рефератов научных текстов в области ИКТ необходима разработка специализированных метрик, учитывающих как синтетический характер русского языка, так и особенности научного дискурса ИКТ.

ЛИТЕРАТУРА

1. BLEU: a method for automatic evaluation of machine translation / K. Papineni, S. Roukos, T. Ward, W. J. Zhu // ACL-2002: 40th Annual meeting of the Association for Computational Linguistics. 2002. P. 311–318.
2. Chin-Yew Lin. ROUGE: A Package for Automatic Evaluation of Summaries // Text Summarization Branches Out. Barcelona : Association for Computational Linguistics. 2004. P. 74–81.

УДК 81 '33

Катермина Вероника Викторовна
доктор филологических наук,
профессор
Кубанский государственный
университет
г. Краснодар, Россия

Veronika Katermina
Doctor of Philological Sciences
Professor
Kuban State University
Krasnodar, Russia
katermina_v@mail.ru

Гнедаш Анна Александровна
кандидат политических наук,
доцент
Кубанский государственный
университет
г. Краснодар, Россия

Anna Gnedash
Candidate of Political Sciences,
Associate Professor
Kuban State University
Krasnodar, Russia
anna_gnedash@inbox.ru

ОПЫТ АНАЛИЗА
БОЛЬШИХ ЛИНГВИСТИЧЕСКИХ ДАННЫХ ПОСРЕДСТВОМ ИИ
(на примере исследования российских ВК-сообществ
в сфере семьи и родительства)

Статья посвящена кейсу применения технологий искусственного интеллекта к анализу больших лингвистических данных российских сообществ «ВКонтакте», тематически ориентированных на семью и родительство. Корпус из ≈ 480 тыс. постов трех крупнейших групп («Родители и педагоги! Наши дети», «Родители в теме», «Я МАМА!™ Дети. Семья») был собран в трех временных точках (февраль 2021, 2022 и 2024 гг.) и подвергнут сквозной обработке на базе ИИ «Концептоскоп»: токенизация, лемматизация, LDA-моделирование тем, кластеризация SVO-связок, оценка полярности (TextBlob-RU) и построение когнитивных графов. Полученные метрики показали, что во всех сообществах семейный дискурс следует общей траектории «ценностный пафос → функциональная прагматика»: героико-патриотические и воспитательные нарративы 2021 г. постепенно вытесняются утилитарными лайфхаками и рекламными рекомендациями 2024 г.; эмоциональная амплитуда сплющивается к нейтральному уровню, а ролевая вертикаль «взрослый наставник/ребенок/герой» растворяется в детоцентричных инструкциях. Эти выводы подтверждают продуктивность ИИ-подхода для фиксации скрытых семантических и аффективных сдвигов и позволяют рассматривать цифровое пространство семьи как сервисно-ориентированный конструкт, управляемый логикой быстрых решений.

Ключевые слова: искусственный интеллект; большие лингвистические данные; социальные сети; семейный дискурс; ВКонтакте; тематическое моделирование; эмоциональный анализ; цифровая социология.

EXPERIENCE OF BIG LINGUISTIC DATA ANALYSIS VIA AI
(A Case Study of Russian VK Communities on Family and Parenthood)

The paper presents a methodological case of applying artificial-intelligence techniques to the analysis of big linguistic data drawn from Russian VKontakte communities focused on family and parenthood. A corpus of approximately 480 thousand posts from three largest groups

(“Parents and Teachers! Our Children,” “Parents in the Know,” and “I MAMA!™ Children. Family”) was compiled for three time-points-February 2021, 2022 and 2024 – and processed end-to-end: tokenisation, lemmatisation, topic modelling with LDA, SVO-clustering, sentiment scoring with TextBlob-RU and construction of cognitive graphs using graph neural networks. Quantitative metrics reveal a common trajectory in all communities from value-laden pathos to functional pragmatics: the heroic-patriotic and educational narratives of 2021 are gradually displaced by utilitarian life-hacks and commercial recommendations by 2024; emotional amplitude flattens toward neutrality, and the hierarchical “adult-mentor/child-hero” role pattern dissolves into child-centred instructions. These findings demonstrate the efficacy of AI-based NLP for detecting latent semantic and affective shifts and suggest that the digital representation of family is increasingly shaped as a service-oriented construct driven by the logic of instant solutions.

Key words: artificial intelligence; big linguistic data; social networks; family discourse; VKontakte; topic modelling; sentiment analysis; digital sociology.

Искусственный интеллект (ИИ), оформленный в качестве своеобразного междисциплинарного фокуса, вбирает в себя достижения компьютерных наук, логики, биологии, политологии, психологии и лингвистики, чем обуславливает способность одной и той же методики одновременно решать инженерные задачи и извлекать глубокие культурно-семиотические смыслы из языковой ткани цифровых коммуникаций [1, с. 270]; именно эта теоретическая всеядность позволяет трактовать ИИ не как вспомогательный инструмент, а как методологический фундамент гуманитарной аналитики больших текстовых массивов.

Благодаря подобной методологической рамке поток неструктурированных сообщений, будь то повседневные статусы, дискуссионные тренды или меметические реплики, трансформируется в иерархизированную совокупность сущностей, отношений и тематических кластеров: процедура автоматического извлечения знаний, укорененная в концепции Natural Language Understanding [2], сегментирует текст, фиксирует лексико-грамматические маркеры, устанавливает логические связи и складывает их в машино-читаемую базу, что, в свою очередь, делает возможной последующую статистико-семиотическую интерпретацию больших корпусов.

Поворотным технологическим моментом становится вовлечение больших языковых моделей (БЯМ), которые, сформированные предварительным обучением на терабайтах разнообразного текста, способны одновременно удерживать синтаксическую целостность, учитывать прагматический контекст и реконструировать скрытые мотивационные паттерны дискурса, тем самым задавая качественно иной стандарт лингвистической аналитики и позволяя без ручной разметки индексировать, категоризовать и эмоционально профилировать миллионы сообщений в режимах, близких к реальному времени [3, с. 25].

Примененные к российским сообществам «ВКонтакте», посвященным семье и родительству, методики Natural Language Processing (NLP, обработка естественного языка) демонстрируют, что масштабная, динамическая и жанрово размноженная речевая среда способна служить чувствительным баро-

метром культурных трансформаций. БЯМ вскрывают скрытые кластеры тем от «цифрового родительства» до пересмотра гендерных ролей, моделируют миграцию ценностных акцентов и позволяют предсказывать какие дискурсивные «узлы напряжения» обострятся в обозримом будущем, превращая лингвистический анализ Больших Данных в инструмент социокультурного прогнозирования.

Объект нашего исследования – цифровой дискурс российских ВК-сообществ, посвященных семье и родительству, то есть совокупность всех постов, опубликованных в трех крупнейших группах ВКонтакте («Родители и педагоги! Наши дети», «Родители в теме», «Я МАМА!™ Дети. Семья»).

Цель исследования – выявить динамику смысловых доминант, ценностных акцентов и эмоциональных регистров семейно-родительского дискурса ВК-сообществ в трех временных точках – февраль 2021 г. (до СВО), февраль 2022 г. (момент начала) и февраль 2024 г. (третья годовщина).

Этапы исследования

1. Сформировать корпус объемом 480 тыс. публикаций путем сплошной выгрузки данных через VK API и экспертной фильтрации ключевых слов «мама», «папа», «дети», «семья» и др.

2. Предобработать тексты: удалить пустые строки, стоп-слова, унифицировать регистр, выполнить лемматизацию.

3. Провести тематическое моделирование методом LDA и SVO-классификации, выделив устойчивые смысловые узлы.

4. Оценить эмоциональную полярность публикаций при помощи TextBlob-RU.

5. Построить когнитивные графы смыслов с использованием комплекса «Концептоскоп» на базе графовых нейросетей.

6. Сопоставить результаты трех временных срезов и интерпретировать тенденции (традиционализм → эмоциональная поддержка → прагматизм).

Исследование базируется на парадигме «Большие лингвистические данные» и когнитивно-лингвистическом подходе к дискурсивному анализу; используется подход Digital Sociology о социальных медиа как барометре общественных изменений; принцип Data-Driven Linguistics, где знания извлекаются напрямую из неструктурированных текстов.

Методы:

- сбор данных через VK API;
- токенизация, лемматизация и стоп-фильтрация на Python;
- тематическое моделирование LDA на матрице CountVectorizer;
- анализ SVO-связок и n-грамм для семантического профилирования;
- расчет эмоциональной полярности TextBlob-RU;
- построение когнитивных графов в «Концептоскоп», работающем на графовых нейронных сетях;
- диахронное сравнение трех корпусов с последующей качественной интерпретацией.

Результаты исследования: как трансформируются семейные репрезентации в трех ВК-сообществах.

«Родители и педагоги! Наши дети», «Родители в теме», «Я МАМА!™ Дети. Семья»

Каждая ВК-площадка прошла собственную, но типологически соотносимую траекторию «ценностный пафос → функциональная прагматика».

В Сообществе «Родители и педагоги! Наши дети» героико-педагогический дискурс 2021 года, насыщенный патриотическими аллюзиями и фигурой взрослого наставника, уже к 2024-у редуцируется до утилитарных чек-листов и детоцентричных лайфхаков. Сообщество «Родители в теме», изначально ориентированное на развернутые методички раннего развития, эволюционирует к форматам микро-подсказок, поддерживая «эмоциональное плато» нулевой полярности и лишь эпизодически добавляя экологические акценты. Сообщество «Я МАМА!™ Дети. Семья», стартовавшее с «папоцентричной» лирики, демонстрирует наиболее резкий кульбит: к 2024 году половина контента превращается в каталог партнерских ссылок и функционирует уже как рекламный хештег в витрине маркетплейса. При этом для всех групп характерно статистически фиксируемое выравнивание эмоционального тонуса к нулю, что можно интерпретировать как аффективную усталость аудитории и переключение коммуникации на режим минимального когнитивного вложения.

Кросс-сопоставление показывает феномен аффективной нивелировки: распределения полярности, какими бы разными они ни были на входе, стремятся к нулю, а спектр эмоций сужается.

Гендерная перспектива подтверждает общий сдвиг от иерархий к растворению ролей. Если в 2021 г. вертикаль «взрослый наставник → ребёнок-герой» задает композицию постов «Родители и педагоги! Наши дети», то к 2024 г. взрослый субъект становится безымянным фасилитатором чужих лайфхаков, а ребенок формально превращается в единственного адресата инструкции. В Сообществе «Я МАМА!™ Дети. Семья» аналогичный процесс идет через стадию «гендерного тандема счастья» (2022 г.) к роли «советчика-покупателя», когда автор говорит читателю, что именно следует класть в корзину ради «гармоничного развития» ребенка. Сообщество «Родители в теме» изначально избегало жестких ролевых оппозиций, однако и здесь в 2024 г. фиксируется переход от индивидуального «родитель-эксперт» к коллективному «мы – команда, улучшающая мир», что эмпирически проявляется в росте запросов об имени или совместных экологических инициативах.

Качественная интерпретация этих сдвигов выводит нас к концепту семьи как сервиса. Когда эмоциональное поле выравнивается, а идеологические знаки изымаются, на первый план выходит прагматика времени и ресурса: быстрее найти совет, схему, ссылку; быстрее получить выгоду; быстрее адаптировать ребенка к бурной внешней среде. Именно этот сервисный вектор задает рост жанрового микротекста: длинная эссеистика стирается в пользу чек-листов и бейдж-постов, а доминирующий медиум смещается к изображению или лаконичному сторис-описанию.

Такое понимание открывает сразу две перспективы. С одной стороны, у гуманитарной науки появляется верифицируемый инструмент описания «сердобольного выхолащивания» ценностей: ИИ-модели позволяют количественно доказать, что престиж героических, патриотических или воспитательных нарративов снижается и замещается сервисной лексикой. С другой – политикам в сфере семейного благополучия и цифрового образования становится яснее, что попытки возвращать в публичный обиход «большой пафос» без учета потребности в быстрой практической выгоде могут наткнуться на барьер аффективного равнодушия.

У всех трех ВК-сообществ эмотивный график сплюснулся почти до нуля. Нейросеть видит: вместо всплесков радости или гордости – ровная линейка информационных сообщений. Это похоже на «аффективную усталость»: людям нужно быстрое решение, а не патетика. Одновременно исчезли героические и патриотические слова; там, где раньше говорили о «подвиге» или «будущем защитнике», теперь сухо объясняют, как успеть сделать поделку до понедельника. Даже мем «папа – герой» выжил лишь в формате рекламного хэштега.

Получается, что в цифровом пространстве семья перестает быть «малой родиной» или «школой гражданина» и превращается в сервисный конструктор: набор быстрых инструкций, скидок и микрорешений. Именно искусственный интеллект раскладывает эту трансформацию по полочкам: от частот слов до диаграмм эмоций и делает видимой тенденцию, которая на глазах выглядела бы как обычная суэта ленты. То есть большие лингвистические данные + ИИ = инструмент, который позволяет заметить, как героический нарратив тихо сдает позиции прагматике, а семейная риторика становится руководством «сделай-сам» в реальном времени.

ЛИТЕРАТУРА

1. Сорокина С. Г. Искусственный интеллект в контексте междисциплинарных исследований языка // Вестник Кемеровского государственного университета. Серия: Гуманитарные и общественные науки. 2023. Т. 7. № 3. С. 267–280. DOI: <https://doi.org/10.21603/2542-1840-2023-7-3-267-280>.
2. Томарева И. Г. Взаимодействие ИИ и когнитивной лингвистики: актуальное положение и будущие направления // Международный научно-исследовательский журнал. 2024. № 6 (144). DOI: <https://doi.org/10.60797/IRJ.2024.144.86>.
3. Громов М. С., Чертовских М. Г. Большие языковые модели: текущее состояние, оценки и прогнозы // Международный бизнес. 2023. № 3 (5). С. 24–31.

УДК 81'33

Красовская Юлия Юрьевна

преподаватель

Белорусский государственный университет

г. Минск, Беларусь

Julija Krasowskaja

Lecturer

Belarusian State University

Minsk, Belarus

hisradaar@gmail.com

ЭФФЕКТИВНОСТЬ СИНТАКСИЧЕСКИ-ОБОГАЩЕННОГО СЕНТИМЕНТ-АНАЛИЗА ДЛЯ РУССКОЯЗЫЧНЫХ И АНГЛОЯЗЫЧНЫХ ТРЕВЕЛ-БЛОГОВ

Данное исследование направлено на повышение точности сентимент-анализа англоязычных и русскоязычных текстов тревел-блогов за счет включения в архитектуру обработки естественного языка синтаксического компонента. Разработана многоступенчатая модель, реализующая автоматическое определение темы текста, отбор релевантных высказываний, синтаксический разбор и аспектно-ориентированный анализ тональности. Сравнительный анализ показал, что интеграция синтаксического анализа позволяет повысить точность, полноту и F1-меру по сравнению с базовыми моделями без синтаксической поддержки. Полученные результаты подтверждают, что учет грамматических зависимостей и аспектной релевантности позволяет улучшить в среднем на 5 % качество автоматической обработки эмоциональной оценки пользовательских высказываний.

Ключевые слова: сентимент-анализ; синтаксический анализ; тревел-блоги; YandexGPT; аспектно-ориентированный анализ; автоматическая обработка текста.

THE EFFICIENCY OF SYNTACTICALLY ENRICHED SENTIMENT ANALYSIS FOR RUSSIAN AND ENGLISH TRAVEL BLOGS

This study aims to improve the accuracy of sentiment analysis of English- and Russian-language travel blog texts by integrating a syntactic component into the natural language processing architecture. A multi-stage model was developed that performs automatic topic detection, selection of relevant utterances, syntactic parsing, and aspect-based sentiment analysis. Comparative evaluation showed that the integration of syntactic analysis increases precision, recall, and F1-score compared to baseline models without syntactic support. The results confirm that accounting for grammatical dependencies and aspectual relevance improves the overall quality of sentiment analysis of user-generated content by an average of 5 %.

Key words: sentiment analysis; syntactic parsing; travel blogs; YandexGPT; aspect-oriented analysis; natural language processing.

В условиях стремительного развития цифровых технологий социальные медиа становятся не только каналом общения, но и важным инструментом поиска информации. Все больше пользователей обращаются к блогам, форумам и платформам с пользовательским контентом при планировании поездок, выборе маршрутов и оценке туристических услуг. Количество пользователей, использующих социальные сети для поиска информации о путешествиях, стабильно растет [1]. Более половины путешественников,

особенно представители поколений Z и миллениалов, предпочитают черпать вдохновение и даже бронировать поездки через Instagram, что делает социальные сети таким же влиятельным ресурсом, как и традиционные туристические медиа. Визуальный формат контента предлагает более живые и достоверные отзывы, чем классические справочники [2].

В свою очередь, автоматическая обработка естественного языка развивается всё стремительнее благодаря прорывам в архитектуре трансформеров, крупномасштабным языковым моделям и мультимодальным подходам. Так, современные трансформеры, включая BERT, GPT-4, демонстрируют способность учитывать контекст и отмечать стилистические особенности анализируемого текста. Исследования последних лет в области обработки естественного языка отмечают активное внедрение мультимодальных моделей, способных обрабатывать текст вместе с изображениями и аудио [3], что открывает возможности для более глубокого анализа пользовательского контента. Одновременно активно развиваются методы few-shot (обучение модели на небольшом количестве предварительно размеченных данных) и zero-shot (обучение модели без предварительно размеченных данных) обучения.

Сентимент-анализ – это одна из ключевых задач автоматической обработки текста, направленная на выявление эмоциональной окраски высказываний. В контексте тревел-блогов он позволяет определить, какие впечатления – положительные, нейтральные или отрицательные – вызывает у пользователей тот или иной объект путешествия: город, достопримечательность, отель, маршрут. Такие данные имеют практическую ценность как для читателей, ищущих рекомендации, так и для индустрии туризма, стремящейся адаптировать предложения под реальные ожидания аудитории. На сегодняшний день автоматический сентимент-анализ чаще всего реализуется с использованием лексиконных подходов, машинного обучения или нейросетевых моделей. Лексиконные методы опираются на заранее составленные словари тональности, тогда как алгоритмы машинного обучения основаны на заранее размеченных корпусах. Нейросетевые архитектуры, особенно трансформеры, позволяют учитывать более широкий контекст высказываний. Однако все эти подходы, особенно при анализе пользовательского контента, сталкиваются с рядом трудностей: ирония, многозначность, нестандартные синтаксические конструкции и стилистическая вариативность часто приводят к ошибочной интерпретации смысла.

Синтаксический анализ представляет собой автоматический метод выявления грамматических структур текста, в ходе которого исходная цепочка токенов преобразуется в деревья составляющих (constituency parsing) или деревья зависимостей (dependency parsing). При создании деревьев составляющих синтаксический анализ производится на основе контекстно-свободной грамматики. Узлы этого древа обозначают как словоформы, так и

грамматические фразы (например, NP, VP). Для построения таких деревьев часто используется преобразование грамматики в нормальную форму Хомского, позволяющее применять эффективные алгоритмы, такие как алгоритм Кока-Янгера-Касами, при этом не теряя выразительности грамматики. Дерево зависимостей, напротив, иллюстрирует отношения «голова – зависимый» между отдельными словами. Узлы соответствуют самим словам, а ребра – типу синтаксической связи (подлежащее, дополнение, модификатор и др.). Этот подход особенно эффективен для языков со свободным порядком слов и задач, где важно понять, какие слова влияют на интерпретацию других.

В контексте сентимент-анализа синтаксическая информация позволяет выявлять точные связи между главными словами и аспектами оценки. Практические исследования показывают, что метод аспектно-ориентированного сентимент анализа на базе библиотеки CoreNLP с деревом зависимостей повысил эффективность сентимент-анализа более чем на 10 % по сравнению с подходами без синтаксического анализа [4].

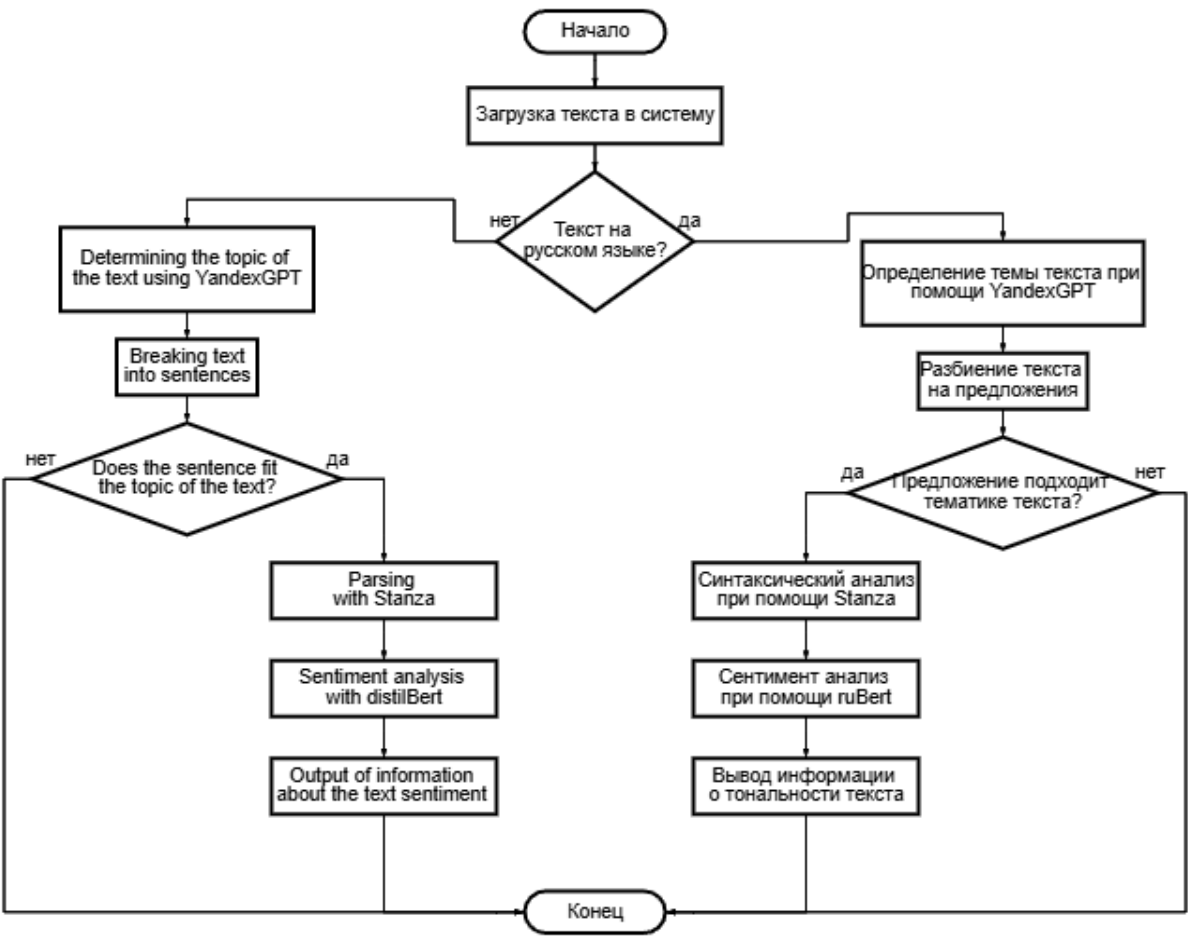
Предварительно нами проведена сравнительная оценка эффективности существующих средств автоматического сентимент-анализа на английском и русском языках [5]. Для эксперимента использовались вручную размеченные тестовые корпуса объемом по 1000 словоформ на русском (URL: <https://clck.ru/3KRP63>) и английском (URL: <https://clck.ru/3KRP2Y>) языках, включающие как положительно, так и отрицательно окрашенные высказывания. Сравнение производилось по основным метрикам (precision, recall, F1-score), а также по способности моделей адекватно определять общую тональность текста. Полученные результаты выявили существенные различия в качестве работы библиотек: нейросетевые модели, такие как RoBERTa, обеспечивали более высокую точность (до 80 % F1-score для русского и английского языков), в то время как лексиконные методы демонстрировали значительно более низкие показатели.

Полученные результаты подтвердили необходимость внедрения структурных лингвистических подходов – прежде всего, синтаксического анализа – для повышения точности анализа тональности исследуемых текстов. В отличие от классических методов сентимент-анализа, которые часто ориентируются только на отдельные маркеры тональности, синтаксический анализ позволяет установить, какое именно слово или группы слов модифицирует определенный объект, и таким образом локализовать эмоционально значимую оценку.

Для проведения данного исследования нами принято решение использовать библиотеку Stanza, разработанную Стэнфордским университетом [7]. Этот инструмент представляет собой современный NLP-фреймворк с открытым исходным кодом, реализующий токенизацию, лемматизацию, морфологический разбор и синтаксический парсинг на основе деревьев зависимостей.

Одним из ключевых преимуществ Stanza является наличие предобученных моделей для более чем 70 языков, включая русский и английский. Stanza использует в своей архитектуре многослойные нейросетевые модели, основанные на би-LSTM и CRF, что обеспечивает высокую точность разметки при сравнительно невысокой вычислительной нагрузке.

Нами разработана многоступенчатая модель, ориентированная на повышение точности сентимент-анализа за счет интеграции синтаксических компонентов. Модель реализует последовательный процесс обработки текстов тревел-блогов. Блок-схема разработанной модели представлена на рисунке.



Блок-схема алгоритма по сентимент-анализу текстов русскоязычных и англоязычных тревел-блогов

В начале работы модели осуществляется загрузка авторизационного ключа сервисного аккаунта в формате JSON, необходимого для аутентификации при обращении к API платформы Яндекс.Облако. С использованием библиотеки `pyjwt` формируется JWT-токен, на основе которого посредством HTTP-запроса к IAM-серверу Яндекс. Облако запрашивается временный IAM-токен, обеспечивающий доступ к модели YandexGPT. Далее к YandexGPT

направляется промт, сформулированный как инструкция к определению темы, на вход которого передается полный текст. В ответ модель генерирует краткое тематическое резюме, которое в дальнейшем используется в качестве семантического фильтра. За этим следует отбор релевантных предложений. Для каждого предложения текста формируется отдельный промт, адресуемый модели YandexGPT, с уточняющим вопросом о тематической релевантности данного высказывания. Ответ модели в бинарной форме (да/нет) служит основанием для включения предложения в дальнейший анализ. Для синтаксической обработки отобранных предложений используется библиотека Stanza, обеспечивающая глубокий лингвистический разбор, включая токенизацию, определение частей речи, лемматизацию и построение деревьев зависимостей. На основе дерева зависимостей извлекаются значимые конструкции, включающие существительные и связанные с ними контекстуальные элементы. Кроме того, аспектно-ориентированный сентимент-анализ для текстов на русском языке основан на модели cointegrated/rubert-tiny-sentiment-balanced, реализованной через интерфейс transformers.pipeline. Для каждого извлеченного контекста производится определение тональности, выраженной в одной из трех категорий: положительной, отрицательной или нейтральной. Далее происходит суммирование результатов: общее распределение эмоциональных категорий по всем релевантным предложениям фиксируется количественно, и на основании доминирующего класса выводится интегральная оценка общей тональности исходного текста. При этом в случае приблизительного равенства между ведущими категориями результирующая тональность маркируется как нейтральная.

В случае обработки англоязычных текстов архитектура модели сохраняет общую структуру, однако адаптируется к специфике языка посредством изменения как лингвистических ресурсов, так и параметров взаимодействия с языковой моделью. В частности, текстовые промты, направляемые на вход модели YandexGPT для определения темы и релевантных предложений, автоматически формируются на английском языке, что позволяет избежать искажения смысла при интерпретации запросов. Для синтаксического анализа используется англоязычная модель библиотеки Stanza. Аспектно-ориентированный сентимент-анализ в англоязычной версии модели реализуется с использованием предобученной модели distilbert-base-uncased-finetuned-sst-2-english, предоставляющей точную классификацию эмоциональной окраски текста в категориях *positive*, *neutral* и *negative*. Программная реализация сохраняет единую методологическую основу для анализа, обеспечивая при этом адаптацию к языковым различиям за счет применения соответствующих языковых моделей и синтаксических парсеров.

Результаты оценки эффективности разработанной модели для сентимент-анализа тестового текста на русском языке представлены в таблице 1.

Т а б л и ц а 1

Оценка эффективности разработанной модели
для sentiment-анализа тестового текста на русском языке

Используемые инструменты	Hugging Face Transformers (XLM-RoBERTa)	YandexGPT + Stanza + ruBert
Precision	72 %	78 %
Recall	90 %	92 %
F1-Score	80 %	82 %
Корректность определения итоговой тональности тестового текста	Да	Да

Анализ полученных данных позволяет сделать вывод о том, что интеграция синтаксического анализа в архитектуру sentiment-анализа тестового текста на русском языке способствует повышению общей эффективности модели по всем основным метрикам. В частности, применение базовой модели на основе Transformers (XLM-RoBERTa) без учета синтаксической структуры обеспечило значение F1 на уровне 80 %, в то время как при использовании расширенного подхода, включающего комбинацию наблюдается рост всех показателей: точность увеличилась до 78 %, полнота до 92 %, а F1-мера – до 82 %. В обоих случаях модели корректно определяли итоговую тональность тестового текста, что свидетельствует об общей функциональной состоятельности обеих архитектур.

Результаты оценки эффективности разработанной модели для sentiment-анализа тестового текста на английском языке представлены в таблице 2.

Т а б л и ц а 2

Оценка эффективности разработанной модели
для sentiment-анализа тестового текста на английском языке

Используемые инструменты	Hugging Face Transformers (RoBERTa)	YandexGPT + Stanza + DistilBert
Precision	83 %	86 %
Recall	78 %	90 %
F1-Score	80 %	88 %
Корректность определения итоговой тональности тестового текста	Да	Да

Результаты анализа англоязычного тестового текста демонстрируют аналогичную тенденцию повышения эффективности модели при включении синтаксического анализа. Базовая модель RoBERTa обеспечила F1 80 %, в то время как интегрированный подход, продемонстрировал улучшенные значения всех метрик: точность возросла до 86 %, полнота – до 90 %, а F1-показатель – до 88 %. Обе модели корректно определили итоговую тональность текста, однако прирост эффективности в расширенной конфигурации подтверждает, что синтаксически обоснованная сегментация текста и аспектно-ориентированная классификация позволяют существенно повысить качество sentiment-анализа текста.

Наше исследование продемонстрировало, что включение синтаксического анализа в архитектуру автоматического sentiment-анализа позволяет в среднем на 5 % повысить точность обработки текстов как на русском, так и на английском языках. Разработанная модель, основанная на многоступенчатом подходе, в котором последовательно реализуются тематическая фильтрация, отбор релевантных высказываний, синтаксический разбор и аспектно-ориентированный анализ, показала устойчивое преимущество по сравнению с базовыми моделями, не учитывающими структурные особенности текста.

ЛИТЕРАТУРА

1. Social Media Usage Approach in Travel // Phocuswright URL: <https://www.phocuswright.com/Travel-Research/Consumer-Trends/Social-Media-Usage-Approach-in-Travel> (date of access: 07.07.2025).
2. Exploring Gen Z and Millennial Travel Habits // Skift Research URL: <https://research.skift.com/reports/exploring-gen-z-and-millennial-travel-habits/> (date of access: 05.07.2025).
3. M. Mohammadi, E. Tajik, R. Martinez-Maldonado, Sh. Sadiq, W. Tomaszewski, H. Khosravi Artificial intelligence in multimodal learning analytics: A systematic literature review // Computers and Education: Artificial Intelligence. 2025. № 8.
4. M. Moayad Aziz, A. Abu Bakar, M. RidzwanYaakubCoreNLP dependency parsing and pattern identification for enhanced opinion mining in aspect-based sentiment analysis // Journal of King Saud University - Computer and Information Sciences. 2024. №36, Issue 4
5. Красовская Ю. Ю. Оценка эффективности Python-библиотек для sentiment-анализа (на материале англоязычных и русскоязычных тревел-блогов) // От слова к дискурсу : материалы Междунар. науч. конф.. Минск: МГЛУ, 2025. С. 180–182.
6. Stanza – A Python NLP Package for Many Human Languages // GitHub URL: <https://stanfordnlp.github.io/stanza/> (date of access: 03.07.2025).

УДК 81'32

Липатова Дарья Андреевна
студент департамента филологии
Научный исследовательский
университет
«Высшая школа экономики»
г. Санкт-Петербург, Россия

Daria Lipatova
Student of the Department of Philology
National Research University «Higher
School of Economics»
Saint Petersburg, Russia
dalipatova@edu.hse.ru

СРАВНЕНИЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ И МАРКОВСКИХ ЦЕПЕЙ ДЛЯ ГЕНЕРАЦИИ АНЕКДОТОВ НА РУССКОМ ЯЗЫКЕ

В статье рассматриваются возможности автоматической генерации анекдотов на русском языке с использованием марковских цепей и больших языковых моделей (large language models, LLM). Целью исследования является сравнительный анализ этих методов в задаче генерации юмористических текстов, учитывающих формальные признаки анекдота. Для обучения марковских моделей был собран корпус из более чем 300 000 анекдотов, очищенный от дубликатов и выбросов. Результаты показали, что увеличение параметра *state_size* улучшает связность и осмысленность текстов, но снижает их разнообразие, при этом многие генерации представляют собой комбинации уже существующих анекдотов. В работе также тестировались LLM: GPT-4, YandexGPT (3.0, 5.0 Lite, 5.0 Pro), Gemini (2.0 Flash, 2.0 Flash Thinking, 2.5 Pro) и Grok 3. Выяснилось, что некоторые из них (особенно GPT-4 и Gemini 2.0 Flash) лучше сохраняют ключевые признаки жанра анекдота, создавая тексты с «речью от автора» и характерными персонажами, тогда как другие отклоняются от этих признаков или выдают заготовленные варианты. Результаты исследования подчеркивают необходимость дальнейших экспериментов с промптами и обучающими данными для повышения оригинальности и качества генерации.

Ключевые слова: обработка естественного языка; автоматическая генерация текста; анекдоты; большие языковые модели; марковские цепи.

COMPARISON OF LARGE LANGUAGE MODELS AND MARKOV CHAINS FOR GENERATING JOKES IN RUSSIAN

The article deals with the possibilities of automatic generation of anecdotes in Russian using Markov chains and large language models (LLM). The aim of the study is a comparative analysis of these methods in the task of generating humorous texts that take into account the formal features of the anecdote. A corpus of more than 300,000 anecdotes, cleaned of duplicates and outliers, was collected to train the Markov models. Results showed that increasing the *state_size* parameter improves the coherence and meaningfulness of the texts but reduces their diversity, with many generations being combinations of pre-existing anecdotes. The LLMs tested in this work are GPT-4, YandexGPT (3.0, 5.0 Lite, 5.0 Pro), Gemini (2.0 Flash, 2.0 Flash Thinking, 2.5 Pro) and Grok 3. It turned out that some of them (especially GPT-4 and Gemini 2.0 Flash) were better at preserving key features of the anecdote genre, producing texts with “speech from the author” and distinctive characters, while others deviated from these features or produced pre-prepared variants. The results of the study emphasize the need for further experimentation with prompts and training data to improve originality and generation quality.

Key words: natural language processing; automatic text generation; anecdotes; large language models; Markov chains.

В условиях автоматизации различных областей человеческой деятельности, включая сферу творчества [1], актуальным становится исследование возможностей генерации юмористических текстов средствами искусственного интеллекта [2; 3]. Анекдоты как лингвокультурный феномен представляют собой сложный объект для генерации, так как их создание требует не только владения языковыми шаблонами, но и учета контекстуальных и когнитивных факторов, лежащих в основе комического эффекта. Однако на текущий момент не было обнаружено примеров систематического анализа генерации анекдотических текстов на русском языке.

Целью этой работы является сравнительный анализ больших языковых моделей (Large Language Models, LLM) и марковских цепей в задаче генерации анекдотов на русском языке. Выбор данных методов обусловлен их принципиальным различием: марковские цепи являются классическим статистическим подходом к генерации текста, в то время как LLM представляют собой современный нейросетевой метод. Марковские цепи основаны на предположении, что вероятность появления следующего элемента текста зависит только от текущего состояния, а не от всей предшествующей последовательности. Большие языковые модели работают на архитектуре трансформеров, которая анализирует текст с помощью механизма внимания, позволяющего находить взаимосвязи и зависимости токенов, тем самым учитывая больший контекст, чем другие модели. Такое сопоставление позволяет провести наглядное сравнение простых и продвинутых алгоритмов генерации.

Для сравнения результатов генерации были выбраны признаки анекдота, выделенные в работе Е. Я. Шмелевой и А. Д. Шмелева [4]. Авторы обращают внимание на наличие постоянного набора персонажей с устойчивыми речевыми и поведенческими характеристиками, известными носителям языка и не требующими представления [4, с. 23]. Кроме того, они акцентируют внимание на структуре текста: разделении на часть «от автора» и речь персонажей [4, с. 31]. Текст «от автора» подчиняется специфическим синтаксическим требованиям: начальные предложения часто имеют не прямой порядок слов, а глаголы преимущественно употребляются в настоящем времени или в форме совершенного вида прошедшего времени с результативным значением [4, с. 32–33].

Материалом исследования стали данные, собранные с сайтов, где пользователи размещают анекдоты (Таблица 1), общим объемом 358408 текстов. После удаления дубликатов, пустых записей и выбросов по длине размер итогового датасета составил 309723 анекдота.

Т а б л и ц а 1

Источники анекдотов для датасета

Адрес ресурса	Подраздел	Кол-во анекдотов
https://anekdotov.net/	«Популярно сегодня» с 1 января 2006 года по 15 февраля 2025 года	36868
https://anekdoty.ru/	Страницы категорий (напр. «про Вовочку»)	6295
https://www.anekdot.ru/	Страница самых популярных анекдотов дня с 8 ноября 1995 года по 15 февраля 2025 года	246598
https://anekdotovstreet.com/	Анекдоты дня с 1 января 2019 года по 15 февраля 2025 года	68647

Сначала рассмотрим результаты, полученные при использовании марковских цепей, для обучения которых применялась библиотека `markovify` [5]. Для генерации был выбран метод `markovify.NewlineText` с параметрами по умолчанию, так как он учитывает переносы строк как границы анекдотов в отличие от `markovify.Text`, который полагается на точки как маркеры конца предложения. Пример генерации:

(1) Так или иначе, но всё равно по ней ботинком, чтобы кому-то понравиться.– А тебе какого размера это вырастет, а уже ссоримся... – Таки я подожду.

Основная проблема результатов – отсутствие смысловой связности: хотя предложения грамматически корректны, их части логически не связаны. Для улучшения результатов был увеличен параметр `state_size` (количество предыдущих слов, учитываемых при генерации следующего) с 2 до 3. Некоторые результаты:

(2) У нас дома все решает папа. А кто у тебя? – Кот. – Какой? – Разжиреловый.

(3) Бредет грустный Жириновский в одиночестве по Виа Монтенаполеоне мимо шикарных миланских бутиков и думает: – Главное, чтобы по утрам меня одевали и причесывали птички, а то разнесу всю деревню к такой-то матери. Ну, в смысле, справа налево. Много думал.

Качество генерации улучшилось: пример (2) демонстрирует полностью связный и смешной текст. В случае (3) осмысленные фрагменты стали длиннее, хотя полной связности достичь не удалось.

При дальнейшем увеличении `state_size` до 4 модель начала иногда возвращать `None` – это происходит, когда невозможно сгенерировать текст, удовлетворяющий условию совпадения с оригиналом менее чем на 70 % (`max_overlap_ratio = 0.7`). Тем не менее, большинство генераций остаются успешными:

(4) Где-то в России. Доктор – пациенту: – Поздравляю. – С чем?! – Я болел

(5) – *Потерпевшая! Назовите наиболее яркую примету преступника? – Он был гонщиком, Нурислам. – Автогонщиком? – Самогонщиком.*

Хотя некоторые сгенерированные предложения остаются нелогичными, доля удачных примеров возросла. Наблюдается закономерность, отмеченная авторами исследования [6]: успешные генерации часто представляют собой комбинации нескольких анекдотов из датасета [6, с. 3]. Например, пример (5) объединяет два анекдота:

1) – *Потерпевшая! Назовите наиболее яркую примету преступника? – Он был без маски.*

2) – *Бабуль, а почему ты в деда влюбилась? – Он был гонщиком, Нурислам. – Автогонщиком? – Самогонщиком.*

Дальнейшее увеличение `state_size` не приводит к существенному улучшению качества, а лишь повышает частоту появления `None` и текстов, практически дословно повторяющих анекдоты из датасета.

Таким образом, анекдотам, сгенерированным марковскими цепями, удастся сохранить особенности данных, на которых они обучались, в частности, диалогическую форму и речь автора в некоторых случаях. Однако часто их сложно назвать оригинальными, так как они часто представляют собой комбинацию из имеющихся в датасете текстов. Это подтверждает ограниченность классических марковских цепей.

Далее обратимся к генерациям, полученным в ходе применения больших языковых моделей. Для того, чтобы иметь возможность выделить общие особенности генераций языковых моделей, мы использовали несколько: GPT-4 [7], YandexGPT [8] (версии 3.0, 5.0 Lite, 5.0 Pro), Gemini [9] (2.0 Flash, 2.0 FlashThinking, 2.5 Pro) и Grok 3 [10]. Во всех случаях генерация производилась по промпту «сгенерируй анекдот», выбранному из-за своей короткости и нейтральности. В некоторых случаях были попытки генерации для других коротких промптов, о них указано отдельно.

Для первых генераций используется модель GPT-4 в интерфейсе ChatGPT от OpenAI. Рассмотрим результаты 10 генераций без ограничений по промпту «сгенерируй анекдот». Все сгенерированные тексты имеют форму диалога, но только в одном из результатов перед диалогом присутствует «речь от автора»:

(6) *Мужчина приходит к психиатру:*

– *Доктор, у меня проблема – я постоянно разговариваю сам с собой.*

– *И вам это мешает?*

– *Нет, но я такой зануда, что даже себя раздражаю!*

Можно заметить, что эта «речь от автора» тоже не совсем отвечает признакам анекдота: хотя в ней используется глагол в настоящем времени, порядок слов остается прямым, а не меняется на обратный.

Получившиеся анекдоты можно разделить на несколько групп по действующим героям. 3 из них можно отнести к категории «у врача»: все из

них представляют собой диалог доктора и пациента, и диалогическая часть начинается с обращения «доктор» (– *Доктор, у меня плохая память...*; – *Доктор, у меня постоянно такое чувство...*). 5 представляют собой диалог мужа и жены, диалог начинается с обращений «дорогая», «дорогой» или «милый». Оставшиеся два – звонок в службу поддержки и разговор с начальником. Можно сделать вывод, что в получившихся анекдотах сохраняется наличие постоянных для анекдотов героев.

Анекдоты получились разными по качеству, но для наиболее удачных из них часто можно найти очень похожие в интернете. Примеры представлены в Таблице 2.

Т а б л и ц а 2

Сравнение некоторых результатов генерации ChatGPT
с анекдотами в открытом доступе

Генерация ChatGPT	Анекдот из интернета
– Дорогой, ты где? – На рыбалке! – А кто это там хихикает? – Рыбы!	Звонок на мобильник: – Милый, ты где? – Я на рыбалке. – А кто это там так громко дышит? – Это щука.
– Доктор, у меня постоянно такое чувство, что меня все игнорируют! – Следующий!	– Доктор, мне кажется, меня все игнорируют! – Следующий!

Эти сходства затрудняют возможность говорить об этих анекдотах как о сгенерированных нейросетью, так как их юмористическая часть была создана человеком.

Перейдем к результатам генерации при помощи YandexGPT от компании «Яндекс». При генерации в базовом режиме (модель YandexGPT версии 3.0) анекдоты часто повторяются и, по своей сути, не являются генерацией – модель просто выдает один из заложенных заранее вариантов. При запросе «придумай анекдот сама» модель отвечает «Я не могу придумывать анекдоты, но могу рассказать вам один», и далее выдает один из имеющихся анекдотов.

При генерации в продвинутом режиме (модель YandexGPT версии 5.0) при запросе «сгенерируй анекдот» модель продолжает выдавать заготовленные анекдоты. Однако при запросе «придумай анекдот сама» начинают появляться уникальные тексты. Несколько примеров:

*(7) Дорогой, я тебе борщ сварила!
А можно сначала посмотреть его резюме?*

(8) Почему программисты всё путают?

Потому что они начинают считать с нуля.

(9) Почему программисты не любят ходить на похороны?

Потому что там нельзя сказать: “Давайте оптимизируем этот процесс!”

По примерам видно, что в сгенерированных текстах есть типичные персонажи, но по другим признакам от анекдота они отличаются. В примере (7) есть речь героев, но нет речи от автора (хотя ее можно несложным образом воссоздать из контекста). Примеры (8) и (9) от анекдота еще дальше, так как в них нет даже прямой речи героев (эти тексты скорее можно назвать короткими шутками). Также можно заметить, что все результаты недлинные и ограничены двумя репликами, что может говорить о некоторой однотипности результатов при генерации без дополнительных условий.

Кроме использованных ранее моделей в интерфейсе Алисы, можно использовать модели YandexGPT 5 Lite и YandexGPT 5 Pro – модели без поиска и других дополнительных функций Алисы. При генерации с Yandex GPT 5 Lite и запросе «придумай анекдот сама» получились тексты, более удачные по выделенным признакам, чем сгенерированные ранее:

(10) Встречаются два математика:

– Ты слышал, у меня кошка научилась считать до трёх!

– Удивительно! А моя собака освоила интегралы. Теперь она считает, что кость длиной в два интеграла лучше, чем кость длиной в один интеграл.

(11) Встретились на улице два филолога. Один говорит:

– Знаете, я сегодня увидел такое редкое и красивое слово, что решил запомнить его навсегда!

– И какое же?

– «Досвидос»! Теперь буду использовать его вместо прощания.

В получившихся анекдотах прослеживается наличие типичных персонажей, есть «речь от автора» и прямая речь героев; «речь от автора» имеет обратный порядок слов, и в ней используются глаголы в настоящем или прошедшем времени. Эти анекдоты уже длиннее, чем генерации предыдущей модели, но у них есть другая частотная особенность: тексты часто начинаются с «встретились/встречаются».

При генерации с YandexGPT 5 Pro получились следующие тексты:

(12) Почему кошка всегда падает на четыре лапы? Потому что у неё есть девять жизней на тренировки!

(13) Почему кролики любят играть в прятки? Потому что у них всегда есть запасное укрытие – собственная нора!

(14) Почему у фотографа всегда хорошее настроение? Потому что он знает: даже самый хмурый день можно превратить в яркий снимок!

Похожие результаты мы получали ранее: такие тексты скорее относятся к шуткам, чем к анекдотам. Здесь также можно обратить внимание на однотипность результатов.

Среди моделей Gemini от компании Google для генерации мы будем использовать 2.0 Flash, 2.0 FlashThinking и 2.5 Pro.

Некоторые примеры генерации с Gemini 2.0 Flash:

(15) Сидят два программиста в баре. Один говорит другому:

– Слушай, я вчера такое обновление установил, у меня теперь кофе сам варится!

Второй отвечает:

– Да ну? А у меня, знаешь, какая программа? Она мне носки стирает и гладит!

Первый с недоверием:

– Не может быть!

Второй:

– Ну, как тебе сказать... Иногда вылетает ошибка «Носки не найдены»...

(16) Встречаются два нейросети. Одна говорит другой:

– Я тут научилась писать стихи, как Пушкин!

Вторая отвечает:

– О, круто! А я могу генерировать картины, как Ван Гог!

Первая:

– Да ладно! А можешь нарисовать мне кота, играющего на пианино?

Вторая:

– Легко! Но... боюсь, у него будет девять пальцев и три хвоста.

Первая:

– Ну, это же нейросеть! Главное, чтобы было креативно!

Можно заметить, что получившиеся тексты длиннее, чем были во всех предыдущих моделях. Также интересно, что кроме наличия «речи от автора» (соответствующей критериям, выделенным ранее) в начале текста, она появляется в течение всего анекдота.

Генерация с «думающей» моделью Gemini 2.0 FlashThinking показала немного другие результаты:

(17) Два друга разговаривают. Один жалуется:

– Мой новый телефон такой умный, что постоянно меня перебивает!

Второй отвечает:

– А мой такой старый, что вообще не слушает!

(18) Приходит пациент к врачу и говорит:

– Доктор, у меня такое чувство, будто я курица!

Врач отвечает:

– Садитесь, больной, сейчас посмотрим ваши... кхм... перья.

Здесь анекдоты получились короче, но все характеристики, как и у результатов работы предыдущей модели, остались прежними.

Наконец, посмотрим на работу модели Gemini 2.5 Pro.

(19) – *Внучек, у меня что-то с компьютером случилось! На экране какая-то картинка висит и ничего не нажимается!*

– *Бабуль, а какая картинка?*

– *Ну, такой цветочек красивый... в горшочке...*

– *Бабушка, это же заставка! Ты просто мышкой подвигай!*

– *Ах, вот оно что! Спасибо, родной! А я уже лейку приготовила, думала, полить его надо!*

(20) *Учительница на уроке математики спрашивает:*

– *Вовочка, сколько будет дважды два?*

Вовочка, не задумываясь:

– *Марь Иванна, а мы покупаем или продаем?*

Здесь появляются примеры разной длины и разных форм: в первом примере нет «речи от автора», но во втором она появляется в нескольких местах.

Последняя модель, которую мы попробуем использовать для генерации, – это Grok 3, разработанный компанией xAI. Ниже представлены несколько результатов генерации:

(21) *Мама спрашивает сына:*

– *Почему ты так долго сидишь за компьютером?*

Сын:

– *Мам, я не сижу, я спасаю мир от пиксельных монстров, это моя миссия!*

(22) *Парень приходит к гадалке и спрашивает:*

– *Скажите, когда я разбогатею?*

Гадалка смотрит в шар и говорит:

– *Через 10 лет, если перестанешь спрашивать глупости и начнешь работать.*

(23) *Учитель спрашивает ученика:*

– *Почему ты не сделал домашку?*

Ученик:

– *Да я вчера весь день пытался понять, почему у кота девять жизней, а у меня ни одной нормальной.*

В результате работы этой модели получились небольшие анекдоты, в которых есть типичные персонажи, «речь от автора», в которой, однако, прямой порядок слов, и прямая речь героев. Стоит отметить сходство с генерациями Gemini, в которых речь автора появляется не только в начале анекдота, а перед каждой репликой.

В этой работе были проанализированы возможности генерации анекдотов с использованием марковских цепей и больших языковых моделей и были получены следующие выводы:

1. Марковские модели сохраняют те признаки анекдота, которые сохранены в примерах обучающей выборки, поэтому количество сохраненных

признаков варьируется. Их генерации часто представляют собой комбинацию больших фрагментов анекдотов из датасета, что может приводить к повышенному комическому эффекту, но понижает их оригинальность.

2. Некоторые LLM (особенно Gemini 2.0 Flash, GPT-4) более последовательно воспроизводят формальные признаки анекдота. Однако другие модели могут выдавать готовые шаблоны или шутки вместо анекдотов. Для некоторых LLM также характерна генерация анекдотов схожей структуры и длины.

В дальнейших исследованиях возможно сравнение генераций LLM при использовании различных промптов, дообучение моделей на более разнообразных и качественных датасетах.

ЛИТЕРАТУРА

1. Ivcevic Z., Grandinetti M. Artificial intelligence as a tool for creativity [Electronic resource] // Journal of Creativity. 2024. V. 34. № 2. DOI: <https://doi.org/10.1016/j.yjoc.2024.100079>
2. Gorenz D., Schwarz N. How funny is ChatGPT? A comparison of human- and A.I.-produced jokes. [Electronic resource] // PLOS ONE 19(7). 2024. DOI: <https://doi.org/10.1371/journal.pone.0305364>
3. Ritchie G. Can Computers Create Humor? // AI Magazine. №30(3). 2009. P. 71–81.
4. Шмелева Е. Я., Шмелев А. Д. Русский анекдот. Текст и речевой жанр. М.: Языки славянской культуры, 2002. 144 с.
5. GitHub – jsvine/markovify: A simple, extensible Markov chain generator. [Electronic resource] URL: <https://github.com/jsvine/markovify> (date of access: 03.02.2025).
6. Chaudhary T., Goel M., Mamidi R. Towards Conversational Humour Analysis and Design. 2021. [Electronic resource] (date of access: 16.01.2025). DOI: <https://doi.org/10.48550/arXiv.2103.00536>
7. GPT-4 | OpenAI [Electronic resource] URL: <https://openai.com/index/gpt-4/> (date of access: 26.04.2025).
8. Нейросеть Алиса для решения любых задач. URL: <https://alice.yandex.ru/> (дата обращения: 28.04.2025).
9. Google Gemini [Electronic resource] URL: <https://gemini.google.com/> (date of access: 29.04.2025).
10. Grok [Electronic resource] URL: <https://grok.com/> (date of access: 29.04.2025).

УДК [811.111+811.161.1]'342.3(045)

Медведева Наталья Георгиевна
аспирантка кафедры фонетики
и практики англоязычной речи
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Natalia Medvedeva
PhD student of English Phonetics
and Speech Practice Department
Belarusian State University
of Foreign Languages
Minsk, Belarus
Enatalia7medvedeva@gmail.com

Яскевич Виталий Валерьевич
кандидат филологических наук,
заведующий кафедрой фонетики
и практики англоязычной речи
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Vitali Yaskevich
PhD in Philology,
Head of English Phonetics and Speech
Practice Department
Belarusian State University
of Foreign Languages
Minsk, Belarus
vitvalyas@gmail.com

ИСПОЛЬЗОВАНИЕ АЙТРЕКИНГА НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ИССЛЕДОВАНИИ ВОСПРИЯТИЯ АККОМОДАЦИОННОГО ВАРЬИРОВАНИЯ АНГЛИЙСКИХ И РУССКИХ СОГЛАСНЫХ

В исследовании используется окулографический метод для анализа восприятия аккомодационного варьирования согласных [p] и [п] в английском и русском языках. Эксперимент с носителями русского языка, владеющими английским, выявил, что артикуляторная близость гласных в дифонах с соответствующими согласными усиливает аллофоническую конкуренцию, увеличивая количество ошибок при распознавании и время идентификации, особенно в английском языке как иностранном.

Ключевые слова: айттрекинг; парадигма «Визуальный мир»; аккомодационное варьирование; перцептивная фонетика; иностранный язык; родной язык.

APPLICATION OF AI-POWERED EYE-TRACKING IN THE STUDY OF THE PERCEPTION OF ENGLISH AND RUSSIAN CONSONANTS ACCOMMODATIVE VARIATION

The study employs eye-tracking to investigate the perception of accommodative variation of the consonants [p] and [п] in English and Russian. Experiments with native Russian speakers proficient in English revealed that articulatory similarity of vowels intensifies allophonic competition, increasing errors and identification time, particularly in English as a foreign language.

Key words: eye-tracking; visual-world paradigm; accommodative variation; auditory phonetics; foreign language; native tongue.

Изучение восприятия речи представляет собой междисциплинарное направление, объединяющее лингвистику и нейронауки. Основой для развития этой области послужила модель Вернике–Гешвинда, предложенная более 150 лет назад немецким невропатологом Карлом Вернике и позднее усовершенствованная американским неврологом Норманом Гешвиндом [1, р. 408]. Эта модель стала отправной точкой для становления таких дисциплин, как нейро- и психолингвистика, заложив фундамент для дальнейших исследований восприятия речи. Развитие данных дисциплин неразрывно связано с появлением передовых цифровых технологий, позволяющих глубже понять нейронные и когнитивные механизмы, лежащие в основе процесса обработки речевой информации. Данные технологии включают в себя различные методы нейровизуализации, применение искусственного интеллекта для моделирования процесса восприятия речи [2], а также окулографию (айтрекинг) – технологию отслеживания движения глаз.

Метод окулографии широко применяется в психолингвистических исследованиях благодаря разнообразию стационарного и мобильного оборудования, а также гибкому программному обеспечению, позволяющему адаптировать эксперименты под широкий спектр исследовательских задач и требований. Впервые применение айтрекера в психолингвистическом исследовании было осуществлено в рамках экспериментального подхода «парадигма визуального мира» (Visual-world paradigm), разработанного Р. М. Купером в 1974 году. Во время проведения эксперимента в данной парадигме, участникам предлагалось смотреть на изображения или слова на мониторе с параллельным прослушиванием речи, в то время как айтрекер фиксировал направление движения глаз, задержку взгляда на представленных объектах, синхронизируя данные наблюдения с воспринимаемыми участниками аудиосигналами [1, р. 458]. В основе данного подхода лежит гипотеза, согласно которой траектория движения глаз и фиксация взгляда на объектах отражают когнитивные процессы, происходящие у испытуемых во время восприятия звучащей речи, что обеспечивает возможность более глубокого изучения механизмов данного процесса [3, р. 455]. Так, в своем эксперименте Р. М. Купер обнаружил, что испытуемые направляют взгляд на изображение льва, когда слышат слово «лев» или его часть, а также фиксируют взгляд на изображениях льва, змеи и зебры, услышав слово «Африка» [4, с. 84]. Также было установлено, что активация движения взгляда в направлении изображения происходила в течение 200 мс с начала воспроизведения слова на аудиозаписи, иногда в пределах звучания первой фонемы слова [5, с. 152], что свидетельствовало о возможности испытуемых предвосхищать слова, которые должны прозвучать на аудиозаписи, основываясь на подсказках из контекста, синтаксической и грамматической структурах фраз, а также акустических характеристиках начальных звуков слов.

Несмотря на значительный потенциал айтрекинга для исследования процессов восприятия речи, данный метод получил широкое признание в психолингвистическом сообществе лишь спустя два десятилетия, когда в 1995 году М. К. Таненхаус провел исследование, посвященное изучению процесса обработки синтаксических структур с использованием окулографии. Таненхаус также поспособствовал началу применения данного метода в фонетических исследованиях. Одним из первых наиболее влиятельных исследований в этой сфере стала совместная работа П. Д. Аллопенны, Дж. С. Магнусона и М. К. Таненхауса, проведенная в 1998 году. В ходе окулографического эксперимента испытуемым предлагалось перемещать изображения на мониторе, следуя инструкциям на параллельно прослушиваемой аудиозаписи, например, «Pick up the beaker; now put it below the diamond» («Возьмите мензурку; теперь поместите ее ниже ромба»), при этом на экране были представлены четыре изображения: *beaker* ‘мензурка’ – так называемый «целевой объект» эксперимента, *beetle* ‘жук’ – «когортный конкурент», имеющий идентичные начальные фонемы с целевым объектом, *speaker* ‘динамик’ – «конкурент по рифме», имеющий идентичную целевому слову финальную часть, и *carriage* ‘коляска’ – «отвлекающий стимул», не связанный фонологически с целевым *beaker*. Полученные данные показали, что фиксация взгляда на целевом объекте *beaker* и когортном конкуренте *beetle* начиналась с момента произнесения целевого слова (поскольку оба объекта начинались на идентичные звуки), но фиксации на конкуренте *beetle* уменьшались по мере произнесения целевого слова. При этом фиксации на изображениях начинались уже спустя 200–300 мс после начала воспроизведения слова, что свидетельствует о быстрой интеграции речевой и визуальной информации. Важно, что конкурент по рифме, несмотря на несовпадение звуков в начале слова, имел более длительные фиксации по сравнению с *carriage*. Полученные результаты свидетельствуют о том, что в процессе восприятия речевого сигнала слушатели одновременно оценивают все слова (изображения), частично соответствующие входной информации до тех пор, пока одно из них не будет иметь достаточно признаков для окончательной идентификации. Авторы исследования также отмечают, что «лексическая конкуренция», подразумевающая одновременную активацию конкурирующих слов вместе с целевым, сохраняется даже после того, как целевой объект становится фонетически различимым, что свидетельствует о вероятностной природе процесса распознавания устной речи. Исследование, однако, не рассматривало влияние акустических и артикуляционных характеристик звуков на лексическую активацию, но, согласно авторам, айтрекинг может быть особенно эффективен в изучении влияния акустической информации на распознавание слов [6, p. 422, 437–438].

Целью нашего пилотного исследования было изучение особенностей восприятия аккомодационного варьирования, а также установление характеристик гласных звуков, повышающих успех распознавания согласных аллофонов в комбинациях согласный+гласный (СГ) в начальном положении в слове. Дополнительно мы также стремились выявить темпоральные особенности процесса перцептивной идентификации, определяя степень когнитивной нагрузки данного процесса. Для реализации поставленных целей был организован окулографический эксперимент, в котором приняли участие десять носителей русского языка, владеющих английским на продвинутом уровне. Участникам предлагалось прослушать изолированные аллофонические реализации английского смычно-взрывного согласного [p], а также русского аллофона [п] в начальной позиции в слове, включающего переходной участок последующего гласного звука, и определить соответствие воспринимаемых звуков объектам на параллельно предъявляемых на экране изображениях, идентифицируя слова, начинающиеся с данных аллофонов. Всего каждую реализацию можно было прослушать 3 раза. Выбор данного согласного звука был обусловлен его сложной акустической структурой, позволяющей проводить разносторонние наблюдения. Материал на русском языке был включен для возможности проведения сопоставительного анализа, так как полагается, что перцептивная идентификация звуков родного языка в данном эксперименте будет проходить намного успешнее, чем в иностранном. Изображения были отобраны с учетом высокой частотности употребления соответствующих им слов в речи. Применение изображений вместо напечатанных слов способствовало упрощению процесса соотнесения акустической и визуальной информации, минимизации когнитивной нагрузки и повышению точности измерений, что обеспечило более надежные экспериментальные результаты. Одновременно с воспроизведением аудиостимулов испытуемым предъявлялись несколько изображений. Соответствующие изображениям лексические единицы были представлены пятью английскими и шестью русскими словами, обладающими как дистантными, так и близкими артикуляционными характеристиками ударных гласных, например в словах: *raw* 'лапа', *pearl* 'жемчужина', *pig* 'свинья', *puppy* 'щенок', *pool* 'бассейн' и *пыль*, *Пэн*, *повар*, *пиво*, *пальма*, *пуговица*.

Для получения окулографических данных использовалась программа «Beam Eye Tracker» версии 1.0 с использованием вебкамеры с частотой 30 кадров в секунду. Данная программа использует алгоритмы машинного обучения (AI-powered eye-tracking), при помощи которых искусственный интеллект распознаёт положение глаз, зрачков и головы, рассчитывая фиксации взгляда, саккады (быстрые движения глаз) и повороты головы. Во время эксперимента проводилась запись экрана с отображением направления движения взгляда в виде круга, который не был виден испытуемым (см. рисунок

ниже). Запись экрана производилась при помощи программы OBS Studio версии 25.0, с последующей ручной покадровой обработкой данных айтрекинга в программе Adobe Premiere Pro версии 24.5.0.

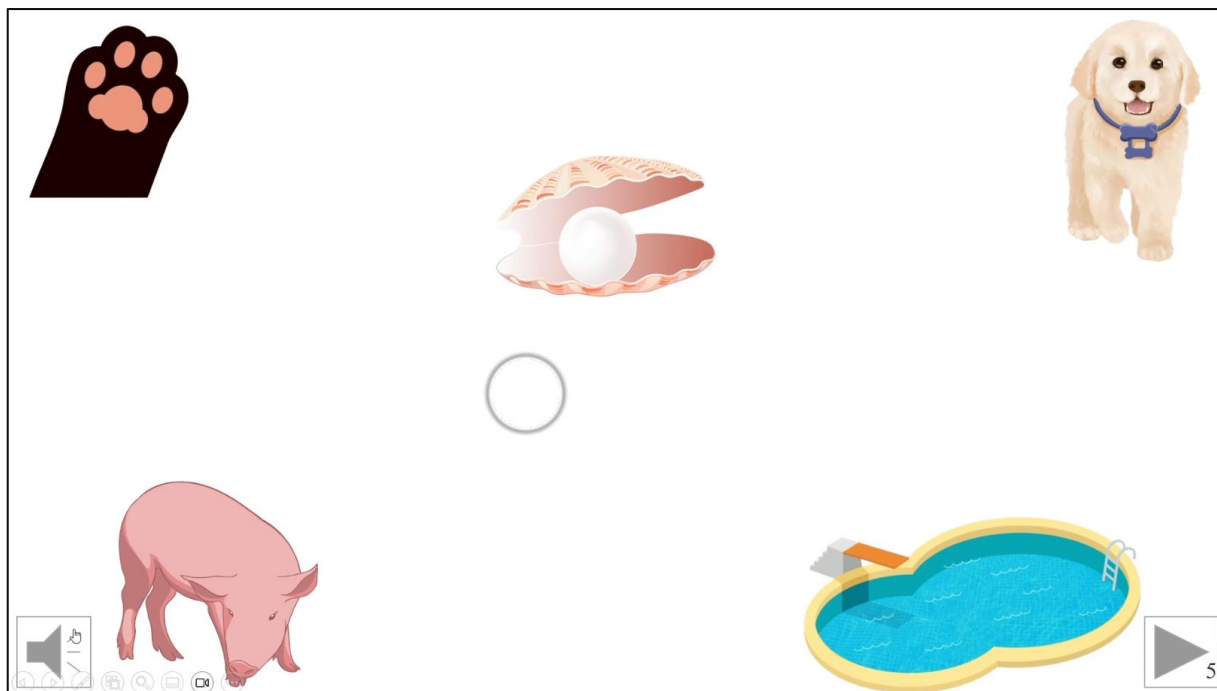


Рис. 1. Пример предъявления визуальных стимулов
и отображения движения взгляда

В результате проведённого эксперимента были определены: средняя длительность фиксации взгляда на каждом визуальном стимуле; общее время, затраченное на принятие решения по каждому слайду с визуальными стимулами (в миллисекундах, от начала первого прослушивания до выбора ответа); количество прослушиваний и ответы, выбранные участниками. Результаты представлены в таблицах 1 и 2 ниже.

Т а б л и ц а 1

Английский язык, процент распознаваний
и конкурирующие аллофоны

Стимул	Процент распознаваний	Конкурирующие аллофоны
pu:	80 %	pɜ:; pi
pɜ:	40 %	pʌ; pɜ:
pʌ	80 %	pɜ:; pɜ:
pɜ:	80 %	pu:

Т а б л и ц а 2

Русский язык, процент распознаваний
и конкурирующие аллофоны

Стимул	Процент распознаваний	Конкурирующие аллофоны
пы	90 %	по
пу	100 %	—
пэ	70 %	пы; па
па	100 %	—
по	80 %	пу
пи	100 %	—

Среднее количество прослушиваний реализаций английского языка составило 2,8 раза, русского – 1,9 раза. Среднее время принятия решения для английских аллофонов достигло 6447 мс, для русских – 3423 мс, с более быстрыми реакциями с примерами из русского языка, как и предполагалось ранее.

Наибольшие трудности в распознавании английских аллофонов вызвали целевые аллофоны слова *pearl* (ошибочно идентифицировано как *puppy* в 50 % случаев, *pearl* – в 40%, *paw* – в 10%) и *pool* (*pool* – 30 %, *pig* – 30 %, *pearl* – 20 %, *paw* – 10 %, *puppy* – 10 %). Ошибочная идентификация *pearl* как *puppy* в 50 % случаев объясняется артикуляторной близостью гласных /з:/ и /ʌ/, которые относятся к среднему подъему и смешанному ряду, различаясь лишь по степени подъема (узкий и широкий соответственно). Неверная идентификация *pool* как *paw* в 10 % случаев, а *paw* как *pool* в 20 % случаев может объясняться близостью гласных звуков по признаку огубленности в рассматриваемых дифонах, а также закрытой артикуляцией звука /ɔ:/ в современной орфоэпической норме. Более продвинутая вперед артикуляция звука /u:/ в современном английском также объясняет его смешение с кратким гласным /ɪ/ в 10 % случаев при распознавании дифона *pu*: [7, с.14].

В части эксперимента, включающего согласные аллофоны русского языка, умеренные трудности возникли при идентификации аллофонов в словах *Пэн* и *повар*. Аккомодационное сочетание /пэ/ в слове *Пэн* было верно распознано в 70 % случаев, тогда как в 20 % и 10 % случаев его ошибочно идентифицировали как /пы/ (*пыль*) и /па/ (*пальма*) соответственно. Эти ошибки обусловлены относительной артикуляторной близостью гласных в данных комбинациях по вертикальному положению языка.

При идентификации аллофонов участники допустили меньше ошибок в русском языке (6) по сравнению с английским (12), что указывает на боль-

шую сложность распознавания иноязычных аллофонов, вероятно, из-за меньшей автоматизации процессов восприятия в иностранном языке. Анализ ошибочных распознаваний в английском языке показал следующее распределение: 50 % ошибок связаны с комбинациями согласный+гласный, где гласные близки по ряду и подъему; 25 % – с гласными, близкими по ряду; 25 % – с гласными, близкими по подъему. Таким образом, близость гласных по ряду и подъему является ключевым фактором ошибок. В русском языке ошибки распределились следующим образом: 83 % связаны с гласными, близкими по ряду, и 17 % – с гласными, близкими по ряду и подъему. Отсутствие ошибок, связанных только с подъемом, указывает на большую надежность этого признака при идентификации аллофонов в родной речи.

Ошибки возникали среди аккомодационных сочетаний с гласными, максимально близкими по своим артикуляционным характеристикам. Это подтверждает, что близость гласных в комбинациях СГ усложняет идентификацию аллофонов независимо от языка, что согласуется с данными Аллопенны и др. о вероятностной активации конкурирующих лексических единиц на основе частичного фонетического совпадения. Подобно их наблюдениям, где фиксации взгляда отражали одновременную активацию когортных и рифмованных конкурентов, наши результаты показывают, что артикуляторная близость гласных в дифонах усиливает конкуренцию между лексическими кандидатами, затрудняя точное распознавание слов.

Участники эксперимента демонстрировали склонность к более продолжительной фиксации взгляда на объектах, которые они затем выбирали. Анализ данных айтрекинга выявил тенденцию к вторичной по длительности фиксации на объектах, названия которых содержали аллофон, близкий по акустическим характеристикам к предъявленному аудиостимулу. Результаты эксперимента приведены в таблицах 3 и 4, где длительность фиксации на объектах с соответствующими аллофонами указана в порядке убывания, а выбранный участниками вариант выделен жирным шрифтом.

Т а б л и ц а 3

Средняя длительность фокуса на объектах
при предъявлении аудиостимула (в мс)

Русский вариант		
пы (0,865)	пэ (0,237)	по (0,237)
пу (0,751)	по (0,324)	пэ (0,269)
пэ (1,295)	па (0,631)	по (0,240)
па (0,691)	по (0,236)	пэ (0,100)
по (0,877)	пи (0,120)	пэ (0,111)
пи (0,776)	пэ (0,328)	пы (0,133)

Т а б л и ц а 4

Средняя длительность фокуса на объектах
при предъявлении аудиостимула (в мс)

Английский вариант		
pu: (1,357)	рз: (0,696)	рэ: (0,550)
рз: (0,916)	рл (0,639)	пу: (0,404)
рл (1,578)	рз: (1,115)	рэ: (0,508)
рэ: (0,669)	рз: (0,335)	пу: (0,308)

При воспроизведении аудиостимула [pu:] участники дольше фиксировали взгляд на объекте с начальным [рз:], гласный которого имеет близость к /u:/ по признаку ряда, тогда как следующая по длительности фиксация на [рэ:] определялась сходством с /з:/ по подъему. Эти гласные также характеризуются относительной близостью по ряду. При предъявлении [рл], ошибочно воспринятого большинством испытуемых как [рз:], фиксации распределялись между указанными сочетаниями с преимуществом в 277 мс для [рз:], что отражает повышенную когнитивную сложность при распознавании аллофона [р] в сочетании с гласными смешанного ряда и среднего подъема. Данные гласные, обладая наибольшей акустической схожестью среди всех стимулов исследования, вызывали частые затруднения при идентификации. Повторное прослушивание [рл] и [рэ:] подтверждало устойчивую вторичную фиксацию на [рз:], указывая на определяющую роль признака ряда в идентификации английского аллофона [р] для данной группы слов.

В случае с русскими аллофоническими сочетаниями в последовательности [пы] – [пэ] – [по] наблюдалась близость гласных по ряду, а также сближение [пэ] и [по] по подъему. В порядке фиксаций [пу] – [по] – [пэ] ключевыми факторами выступили ряд и огубленность. При прослушивании [пэ] вторичные фиксации приходились на [па] и [по], которые схожи преимущественно по подъему и в меньшей степени по ряду. Аллофоны [па] – [по] – [пэ] также демонстрировали большую близость по подъему. В последовательности [пи] – [пэ] – [пы] гласные были относительно равноудалены, при этом /и/ и /ы/ ближе по подъему, а /э/ – по ряду. В случае русских аллофонов основными критериями идентификации аллофонического варьирования выступали ряд и подъем, тогда как огубленность оказывала влияние в редких случаях.

Полученные результаты демонстрируют способность айтрекинга выявить перцептивную чувствительность испытуемых к тонким фонетическим

различиям, таким как аккомодационное варьирование звуков, что делает его эффективным инструментом для изучения процессов восприятия речи, как отмечали П. Д. Аллопенна, Дж. С. Магнусон и М. К. Таненхаус. Кроме того, выявленные закономерности указывают на необходимость учета артикуляторной близости гласных при разработке моделей распознавания речи и методик обучения иностранным языкам. В частности, повышенная когнитивная нагрузка при восприятии иноязычных аллофонов (на 88 % больше времени в английском варианте) подчеркивает важность тренировки фонетической чувствительности для улучшения перцептивной идентификации в иностранном языке. Дальнейшие исследования могут быть направлены на изучение влияния других фонетических факторов на аллофоническую конкуренцию в момент слухового восприятия, а также на расширение выборки для повышения статистической надежности данных.

ЛИТЕРАТУРА

1. Knight R. A., Setter J. The Cambridge Handbook of Phonetics. Cambridge : Cambridge Univ. Press, 2022. 720 p.
2. Дмитриев В. Я., Игнатъева Т. А., Пилявский В. П. Развитие искусственного интеллекта и перспективы его применения // Экономика и управление. 2021. Т. 27. № 2. С. 132–138.
3. An Eye-tracking Technique– Visual-World Paradigm / G. Guan, A. Hu, L. Guo, H. Yu // Proceedings of the 2019 3rd International Conference on Economic Development and Education Management (ICEDEM 2019) : Proceedings of the 2019 3rd International Conference on Economic Development and Education Management (ICEDEM 2019). 2019. P. 455–457. DOI: 10.2991/icedem-19.2019.108
4. Cooper R. M. The control of eye fixation by the meaning of spoken language // Cognitive Psychology. 1974. Vol. 6, № 1. P. 84–107.
5. Huettig F., Rommers J., Meyer A. S. Using the visual world paradigm to study language processing: A review and critical evaluation // Acta Psychologica. 2011. Vol. 137, № 2. P. 151–171. DOI: 10.1016/j.actpsy.2010.11.003
6. Allopenna P. D., Magnuson J. S., Tanenhaus M. K. Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models // Journal of Memory and Language. 1998. Vol. 38, № 4. P. 419–439.
7. Ефимова Е. В. Вариативность английских гласных в акцентной структуре фразы (на материале спонтанной речи) : дис. ... канд. филол. наук : 10.02.04. Минск, 2017. 232 л.

УДК 811.581'37+811.581'35

Михалькова Надежда Васильевна

кандидат филологических наук, доцент,
доцент кафедры теории и практики
китайского языка

Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Nadezhda Mikhalkova

PhD in Philology, Associate Professor,
Associate Professor of the Department
of Theory and Practice of the Chinese
Language

Belarusian State University
of Foreign Languages
Minsk, Belarus
nadezhdakr@yandex.ru

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ПРИ ОТБОРЕ РЕПРЕЗЕНТАТИВНОЙ ЧАСТИ МАССИВА ДАННЫХ В ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

В статье предлагается пример авторской методики отбора репрезентативной части большого массива лингвистических данных для проведения семантических исследований. Выявляются этапы анализа, необходимые для проведения выборки материала, определяются математические формулы, на которых может быть основан отбор репрезентативной части материала исследования.

Ключевые слова: математическая статистика; массив данных; репрезентативность; лингвистический анализ; выборка.

USING METHODS OF MATHEMATICAL STATISTICS WHEN SELECTING A REPRESENTATIVE PART OF A DATA ARRAY IN LINGUISTIC RESEARCH

The article offers an example of the author's methodology for selecting a representative part of a large array of linguistic data for conducting semantic research. The stages of analysis necessary for conducting a selection of material are identified, and mathematical formulas are determined on which the selection of a representative part of the research material can be based.

Key words: mathematical statistics; data array; representativeness; linguistic analysis; sample.

Отбор материала лингвистических исследований, зависящий, в первую очередь, от поставленных задач, достаточно часто представляет собой сложный процесс определения той части большого массива языковых данных, которая будет являться наиболее репрезентативной областью единиц. Поскольку обработать огромные языковые ресурсы зачастую не представляется возможным ввиду ограничения человеческих возможностей отдельно взятого автора или коллектива авторов, встает вопрос о выборке определенной сферы языковых знаков, которая, несмотря на неполный объем от существующего в языке, будет содержать закономерности, экстраполируемые на весь массив данных.

Таким большим массивом данных для нашего исследования послужил список всех иероглифических знаков китайской письменности, общее число которых насчитывает 79728 единиц [1]. Понимая, что анализ всех представленных единиц является нецелесообразным ввиду уже выявленных на тот период работы в ходе пилотного исследования повторяющихся на каждом новом отрезке материала закономерностей, стала очевидной необходимость разработки методики, которая бы позволила отобрать репрезентативную часть всех иероглифических знаков китайского языка. В то же время, поскольку исследование базируется на выявлении семантических закономерностей, несомненно значимым стало соединение нескольких показателей, как количественных, так и качественных методики отбора материала исследования.

Решение поставленных в исследовании задач выявления семантических связей детерминативов иероглифических знаков китайского языка, а также определения на этом основании принципов их выбора для рассматриваемых сложносоставных единиц китайской письменности предполагало проведение двунаправленного анализа, предусматривающего,

1) с одной стороны, проспективный подход, т.е. «движение от знака к смыслу», соответственно, от семантики детерминатива к семантике производных от него иероглифов, входящих в разные группы в зависимости от детерминатива,

2) с другой стороны, ретроспективный подход, т.е. «от смысла к знаку», соответственно, от семантической области, объединяющей сложносоставные иероглифы китайского языка, к анализу детерминативов, которые входят в состав данных иероглифов.

С целью отбора материала исследования нами была разработана собственная методика, основанная на методах математической статистики и, соответственно, выделении трех критериев ограничения представленного в лексикографических источниках огромного корпуса единиц: *семантическая диверсификация; продуктивность и репрезентативность*. Предложенная методика представлена и осуществлена впервые и может быть применена в будущем для иных лингвистических исследований, особенно в области корпусного анализа больших массивов данных и искусственного интеллекта.

Наличие критерия *семантической диверсификации* вызвано необходимостью анализа различных семантических типов детерминативов сложных иероглифических знаков китайского языка с целью избежания семантической одноаспектности исследования, поскольку исключительно количественный путь (например, частотность) не позволил бы учесть разные семантические группы детерминативов, где потенциально могли быть зафиксированы иные закономерности. Для выявления семантических типов детерминативов нами был установлен их исходный количественный состав и проведена семантическая классификация, базирующаяся на онтологических свойствах обозначаемых детерминативами понятий, из которой впоследствии извлекались данные для показателя *семантической диверсификации*.

Критерий *продуктивность* предполагал анализ числа вхождений детерминатива в сложные иероглифы китайской письменности, на основании чего нами впервые были выделены высоко-, средне- и малопродуктивные детерминативы. Далее показатели *семантической диверсификации* и *продуктивности* сопоставлялись как относительно всей системы детерминативов, так и отдельных выделенных нами в семантической классификации макро- и микрогрупп через критерий *репрезентативности*.

Репрезентативность представляет собой анализ отношения показателя средней величины ($\bar{X}$) в семантической макрогруппе или микрогруппе детерминативов к общему числу производных сложных иероглифов (одновременное взаимодействие двух критериев: *продуктивности* и *семантической диверсификации*) с целью отбора наиболее репрезентативных семантических групп.

Следовательно, первой задачей явилось определение исходного состава детерминативов китайской письменности, который был впоследствии взят за основу анализа.

Согласно статистической науке, чтобы оценить любое явление, не обязательно исследовать всю генеральную совокупность. Возможность экстраполировать выводы о части на целую единицу доказывается математикой. В частности, П. Л. Чебышевым сформулирован «Закон больших чисел» [2], который гласит, что количественные закономерности массовых явлений проявляются только при достаточном числе наблюдений. Следовательно, с одной стороны, чем больше выборка, тем яснее проявляется общая тенденция и компенсируются случайные отклонения.

Вместе с тем, вопрос о том, как определить предел выборки, насколько количественно большой, соответственно, репрезентативной она будет являться и каким образом эту часть исчислить может быть решен с помощью разработанных математических теорем, а также статистических уравнений, которые стали фундаментом для создания формул расчета вероятности ошибки и размера выборки материала исследования для достижения заданной точности.

В основу расчета выборки детерминативов иероглифических знаков китайского языка нами положены методы математической статистики, с помощью которых осуществляется сбор и обработка статистических данных для получения научных и практических выводов. На практике сплошное исследование (каждого объекта из интересующей нас совокупности) проводят крайне редко. К тому же, если эта совокупность содержит большое число объектов или исследование объекта требует нарушения его функционального стандарта, то сплошное исследование нереально. В таких случаях из всей совокупности случайно отбирают ограниченное число объектов и подвергают их исследованию.

При проведении исследования получают определенные данные, после применения к которым шкалы измерений они становятся числовыми переменными (зависимыми и независимыми). Совокупность элементов, на кото-

рой проводится исследование и которая характеризует все множество (генеральную совокупность), является выборкой. При этом выборка подбирается так, чтобы она представляла все существенные признаки генеральной совокупности, то есть была репрезентативной.

Различают два вида способов отбора: без расчленения генеральной совокупности на части и с расчленением. К первому виду относятся простые случайные отборы (повторные либо бесповторные), когда объекты извлекают по одному из генеральной совокупности. Второй способ отбора включает в себя следующие разновидности соответственно способам расчленения генеральной совокупности. Отбор, при котором объекты отбираются из каждой «типической» части генеральной совокупности, называется типическим.

Если генеральную совокупность делят на число групп, равное объему выборки, с последующим отбором из каждой группы по одному объекту, то такой отбор называется механическим. На практике часто употребляется комбинирование перечисленных способов отбора. Конкретная комбинация способов отбора объектов из генеральной совокупности определяется требованием репрезентативности выборки.

Для того, чтобы наша выборка была репрезентативной, построим выборочное (упорядоченное) распределение величин – детерминативов иероглифических знаков китайской письменности ($n=264^1$) от наибольшей к наименьшей, сопровождая данными об их продуктивности.

Рассчитаем выборочное среднее по формуле:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k, \text{ где: } \bar{x} - \text{выборочная средняя величина; } n - \text{количество}$$

величин (детерминативы иероглифических знаков китайской письменности);

x_k – частотность значения показателей. Выражение $\sum_{k=1}^n x_k$ соответственно,

означает сумму всех X с индексом k от 1 до n и в нашем случае равно 79728 [1].

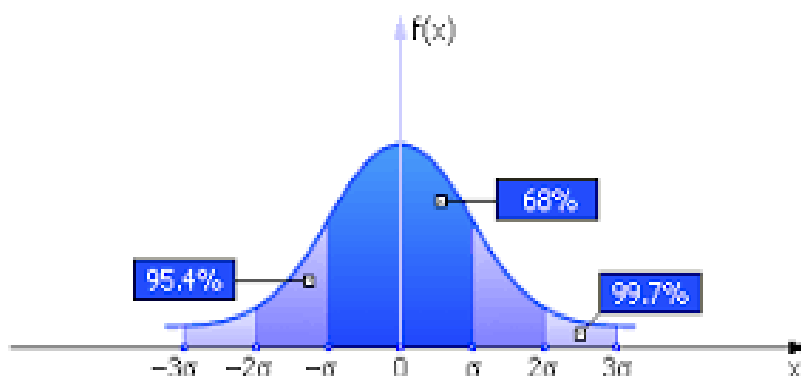
Таким образом, $\bar{x} = 79728/264 = 302$. Для того, чтобы определить меру того, насколько разбросан набор данных, вычислим стандартное отклонение. Стандартное отклонение соответствует квадратному корню из дисперсии и, наряду с дисперсией, является одной из наиболее часто используемых мер

вариабельности признака: $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$

¹ Общее число 266 было сокращено до 264, так как для двух детерминативов в используемом иероглифическом словаре полностью отсутствовали данные относительно сложных иероглифов, куда они входят.

Деление суммы квадратов на число степеней свободы $n - 1$ позволяет сравнивать между собой совокупности, различные по объему. Считается, что дисперсия – более мощный статистический критерий, нежели среднее отклонение, так как больший вклад в дисперсию дают те значения признака, которые расположены дальше от среднего (вклад каждого значения в дисперсию возрастает пропорционально квадрату отклонения от среднего). Стандартное отклонение рассчитаем с помощью приложения Excel. Стандартное отклонение (σ)=499, т.е. разброс наших величин настолько велик, что невозможно определить доверительный интервал, который показывает, в каком диапазоне располагаются результаты выборочных наблюдений.

Согласно правилу трех сигм, данную генеральную совокупность необходимо было бы разделить, удалив часть детерминативов с максимальными и минимальными показателями (рисунок), приблизив выборочный интервал к среднему показателю. Это привело бы к значительным погрешностям выборки ввиду высокого стандартного отклонения. Следовательно, на данном этапе общее число детерминативов невозможно уменьшить, используя только количественный параметр продуктивности.



Нормальное распределение случайной величины

Таким образом, отбор иероглифических данных, согласно критерию – *продуктивности*, показал: наряду с количественной составляющей, необходимо вводить также дополнительный показатель – качественную сторону анализа детерминативов.

Критерий *семантической диверсификации* позволил учесть подобные особенности и максимально охватить разные семантические группы единиц с целью определения как общих для всей системы китайской письменности структурно-семантических закономерностей, так и отличительных свойств отдельных семантических групп детерминативов.

Специфика семантической организация системы детерминативов китайской письменности, а также многоаспектность и разнородность этой системы

единиц подтвердили высказанное ранее предположение о том, что репрезентативная выборка детерминативов должна быть основана не только на количественном, но и на качественном – семантическом параметре, т.е. включать взаимодействие двух критериев: *продуктивности* и *семантической диверсификации*.

Анализ семантических групп детерминативов китайской иероглифической письменности, их продуктивности и деривационного потенциала показал, что существующая широкая диверсификация смысловых компонентов с большой вероятностью может обуславливать наличие различных (в зависимости от семантики детерминатива) закономерностей выбора детерминатива для включения в сложносоставные иероглифы.

Следовательно, отбор детерминативов должен быть осуществлен в соответствии с семантическим критерием, который позволит максимально исключить девиации искомых тенденций.

В каждой из выявленных семантических групп и подгрупп детерминативов были выделены отдельные единицы согласно следующим критериям и процедурам.

Поскольку первый этап расчета выборки позволил определить выборочную среднюю величину по генеральной совокупности, из общей системы нами были исключены детерминативы, показатель *продуктивности* которых не превышал 302 единицы.

Это позволило на втором этапе подсчета ранжировать семантические группы и подгруппы относительно данной величины и не рассматривать семантические подгруппы детерминативов, показатель выборочной средней величины¹ которых не достигал установленного методами математической статистики числа.

Соответственно, в наиболее репрезентативную выборку вошли следующие 12 семантических подгрупп детерминативов: знаки жилищ и их частей ($\bar{X}=389$); знаки природных объектов ($\bar{X}=428$); соматизмы ($\bar{X}=373$); знаки лиц ($\bar{X}=407$); фитонимы ($\bar{X}=529$); зоонимы ($\bar{X}=368$); знаки состояний ($\bar{X}=909$); знаки характеристик ($\bar{X}=322$); знаки видов жидкостей и их состояний ($\bar{X}=599$); знаки огня ($\bar{X}=842$); знаки твердых веществ ($\bar{X}=519$); знаки сыпучих веществ ($\bar{X}=433$).

Отобранные нами на этом этапе семантические подгруппы содержат 103 детерминатива, что составляет 39 % от всего числа детерминативов в китайской письменной системе. При этом общее число производных сложносоставных иероглифов китайской письменности от данных детерминативов – 55560 единиц, репрезентирующих 69,45 % от всего количества иероглифов в китайском языке.

¹ Выборочная средняя величина подсчитывалась по каждой семантической подгруппе путем поиска среднего арифметического на основании показателей продуктивности детерминативов, входящих в данную подгруппу.

Репрезентативность рассчитывалась исходя из того, что средняя величина подгруппы не должна превышать среднюю величину всех подгрупп – $(311,2+880,1)/2=595,65$, при этом, например, семантическая подгруппа «Номинации огня» в расчет не вошла, поскольку состоит из 1 детерминатива, учитывая графическую вариативность. По степени репрезентативности могут быть выделены семантические подгруппы, в которых:

1) совпадают или близки по количественному значению показатели средней величины ($\bar{X}$) и общего числа производных сложных иероглифов (например, семантические подгруппы «Соматизмы», «Фитонимы», «Знаки природных объектов» и др.),

2) существуют значительные расхождения показателей средней величины ($\bar{X}$) и общего числа производных сложных иероглифов от каждого детерминатива в данной семантической подгруппе (например, семантическая подгруппа «Знаки сыпучих веществ»).

В отдельных семантических группах имели место детерминативы с показателями максимальной и минимальной продуктивности, например, «Знаки сыпучих веществ»¹, что привело к необходимости проведения дополнительной итоговой процедуры отбора материала исследования с целью исключения детерминативов с минимальной продуктивностью в каждой семантической подгруппе.

Таким образом, в результате применения первых двух критериев из всей системы детерминативов нами были отобраны 12 семантических подгрупп. Поскольку степень репрезентативности детерминативов данных подгрупп является различной, нами был введен показатель выборочного среднего относительно каждой подгруппы, который позволил исключить наименее репрезентативные детерминативы в каждой из подгрупп.

Следовательно, итоговый расчет выборки детерминативов включает следующие единицы (таблица).

Итоговый состав детерминативов
для осуществления анализа сложных иероглифических знаков,
в которые они входят в качестве компонентов

Название группы	Название подгруппы	Детерминативы	Количество	ПД
Объекты неживой природы	Жилища и их части	宀 ‘крыша’ 广 ‘убежище’ 穴 ‘пещера’ 門/门 ‘ворота’	4(5) ²	689 564 483 616

¹ Детерминатив 𩇛 ‘соль’ имеет показатель частотности – 1, 土 ‘земля’ – 1592.

² В скобках нами указывается число, учитывающее графическую вариативность.

	Природные объекты	日 ‘солнце’ 山 ‘гора’ 月 ‘луна’ 貝/贝 ‘раковина’ 冢 ‘холм’	5(6)	1059 1312 1433 634 1268
	Соматизмы	目 ‘глаз’ 手/手 ‘рука’ 口 ‘рот’ 足 ‘нога’ 心/心 ‘сердце’	5(7)	1073 2280 2854 1050 1986
Объекты живой природы	Лица	人/人 ‘человек’ 女 ‘женщина’ 王 ‘государь’	3(4)	1781 1540 1111
	Фитонимы	木 ‘дерево’ 竹/竹 ‘бамбук’ 艹 ‘трава’	3(4)	2858 1606 3316
	Зоонимы	虫 ‘червь’ 牛/牛 ‘бык’ 隹 ‘короткохвостаяптица’	3(4)	1631 426 332
Обозначения абстрактных сущностей	Состояния	疒 ‘болезнь’	1	909
	Характеристики	饣/食 ‘пища’	1(2)	692
Обозначения веществ	Жидкости	氵/水 ‘вода’	1(2)	2942
	Огонь	火/火 ‘огонь’	1(2)	1685
	Твердые вещества	石 ‘камень’	1	1024
	Сыпучие вещества	土 ‘земля’	1	1592
ИТОГО			29(39)	40746

Таким образом, из всей системы детерминативов китайского языка общее число отобранных для анализа детерминативов составило 39 единиц (включая графическую вариативность), образующих 40746 из 79728 иероглифических знаков китайского языка (51,1 %), что составило репрезентативную часть выборки материала исследования.

ЛИТЕРАТУРА

1. 汉字字典 (Ханыцзыцзидянь). – URL: <https://www.zdic.net/>. (Дата обращения: 10.03.2022).
2. Нифонтов Н. С. Закон больших чисел и теорема Чебышева [Электронный ресурс] / Н. С. Нифонтов, Е. Ю. Маслова // Вестник Академии знаний. – 2017. – № 20 (1). – Режим доступа: <https://cyberleninka.ru/article/n/zakon-bolshih-chisel-i-teorema-chebysheva/viewer>. – Дата доступа: 01.04.2023.

УДК 811.112.2'367.633(043.2)

Мичугина Светлана Викторовна

кандидат филологических наук,

доцент

ГАОУ ВО МГПУ

г. Москва, Россия

Svetlana Michugina

PhD in Philology,

Associate Professor

Moscow City University

Moscow, Russia

michuginasv@mgpu.ru

ВОЗМОЖНОСТИ ИИ ДЛЯ ИССЛЕДОВАНИЯ СЕМАНТИКИ ТЕРМИНОВ ЦВЕТА

В статье представлены результаты сравнительного анализа использования искусственного интеллекта как современного инструмента исследовательской деятельности на примере возможностей двух известных нейросетей, чата ruGPT и чата Qwen, при изучении семантической структуры слова. В качестве материала исследования используются термины цвета в английском языке. В частности, рассматриваются символические значения термина red как одного из базовых цветоименований. Сравнению подвергаются данные лексикографических источников и генерированные результаты нейросетей. Специфика проводимого семантического анализа заключается в использовании лингвистических возможностей анализируемых чатов. Также проиллюстрированы преимущества каждого из ИИ инструментов и сопоставлены выявленные проблемные зоны для проведения лингвистического исследования. Результаты проведенного сравнительного анализа показывают, что чаты ruGPT и Qwen имеют различную эффективность для семантических изысканий, при этом чат Qwen является наиболее перспективным для такого рода исследований, т. к. обладает некоторыми возможностями семантического анализа, что совершенно не свойственно чату ruGPT.

Ключевые слова: искусственный интеллект; сравнительный анализ; чат ruGPT; чат Qwen; лингвистическое исследование; термины цвета.

AI IN SEMANTIC RESEARCH OF COLOUR TERMS

The article presents the results of a comparative analysis of applying artificial intelligence as a modern tool for scientific research, demonstrated with the capabilities of two popular neural networks, that is the ruGPT Chat and Qwen Chat, in studying a semantic structure of a word. The research material represents color terms in the English language. Specifically, the symbolic meanings of the English red, as one of the basic color terms, are examined. The specific nature of the semantic analysis conducted lies in using linguistic appliance of the chatbots analyzed. Besides, the article illustrates the advantages of each of the AI tools, and compares the problem areas identified during the linguistic research. The given results of the comparative analysis show that the ruGPT and Qwen Chats have a varying degree of their effectiveness for semantic inquiries, with the Qwen Chat being more promising for this type of research, as it possesses a certain semantic analysis capability which is entirely absent in ruGPT Chat.

Key words: artificial intelligence; a comparative analysis; ruGPT Chat; Qwen Chat; linguistic research; colour terms.

Информационные технологии занимают все более значимое место в современных лингвистических исследованиях. Искусственный интеллект становится наиболее востребованным цифровым инструментом для получения теоретических и эмпирических данных. Исследователей привлекают, казалось бы, неограниченные возможности искусственного интеллекта. Единственным условием, по мнению многих исследователей, является правильно составленный запрос или промт. Однако, использование ИИ для различных целей имеет свои особенности. Наиболее продуктивным вариантом применения всевозможного рода ИИ чатов является генерация различного рода текстов или их обработка. Возникает вопрос о возможности использования ИИ для исследовательских задач при сборе теоретического материала или получении эмпирических данных в рамках проводимого лингвистического анализа.

Примером популярных ИИ инструментов на данном развитии информационных технологий являются GPT чаты во всех своих модификациях в том числе ruGPT как один из русифицированных вариантов, созданных на базе моделей GPT-4o и GPT-4o-mini для генерации текста, помощи в написании кода и поддержке учебных процессов (Мичугина, 2025). Его соперником в области набирающих популярность нейросетей является чат Qwen, также называемая Tongyi Qianwen, входит в семейство больших языковых моделей, разработанных компанией Alibaba. Последняя версия Qwen 3-235B-A22B выпущена в апреле 2025 года. Согласно открытым сетевым данным, чат Qwen3 превосходит возможности ChatGPT o3 по точности ответов и количеству генерируемых ошибок. По сравнению с GPT-4o Qwen3 показывает сравнимые результаты в программировании и математике, но уступает по способности работать с изображениями, аудио и видео.

Задача нашего исследования заключается в сопоставлении возможности чата ruGPT и Qwen3 в лингвистических исследованиях по изучению семантики прилагательных цвета на примере термина red в английском языке. Остановимся более подробно на символических значениях, т. к. они вызывают наибольший исследовательский интерес. Лексикографическое описание термина red позволяет установить пять значений, только два из которых являются символическими и построены на ассоциации цвета с политическим движением коммунистов, в частности в СССР, и цветом республиканской партии в США (Merriam-Webster Dictionary).

При обращении к чату ruGPT (дата обращения 15.07.2025) с запросом «символические значения прилагательного red в английском языке, получаем семь значений (рис. 1).

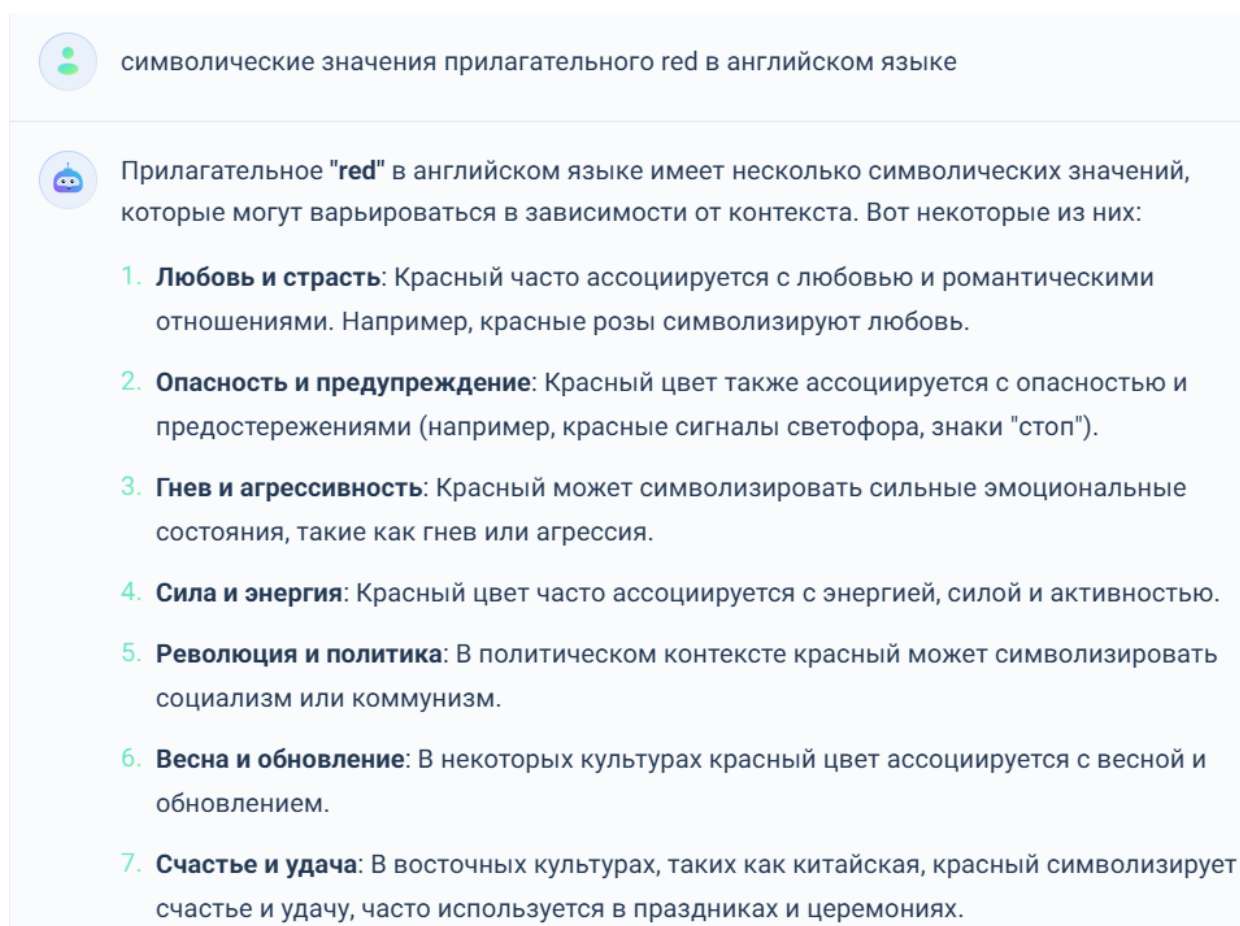


Рис. 1. Данные чата guGPT

Однако значение 3 и 4 не являются символическими, т. к. основаны на метонимическом переносе красного цвета, как правило, лица при проявлении агрессии или здорового румянца, который свидетельствует о силе и энергии человека. Они также зафиксированы и в словаре в семантической структуре прилагательного red.

Лингвистический интерес вызывает значение 1. Оно построено на прямой ассоциации традиционного букета красных роз, который является типичным слотом фрейма «романтическая встреча\свидание». Остается открытым вопрос о возможных ассоциациях с другими сортами цветов и их цветом. Также интересны значения 2, 6 и 7. Первое из них – «опасность и предупреждение» является традиционно выделяемыми в лингвистических исследованиях по семантике цвета (Мичугина, 2005). Значение 6 действительно является символическим, однако требует последующего уточненного запроса, т. к. guGPT не указывает, в каких именно культурах возникла данная ассоциация и существует ли она на настоящий момент. В отличие от значения 7, которое достаточно детализировано и не нуждается в дальнейших уточнениях.

В целом, представленная чатом *gpt* семантическая структура символических значений может быть использована как основа лингвистического анализа, т. к. содержит достоверную информацию, зафиксированную в лексикографических изданиях, что отражено и в других лингвистических исследованиях (Мичугина 2016, 2017). Однако часть из генерированных значений требует дальнейшей детализации, что усложняет алгоритм изучения и анализа семантики изучаемой единицы. Также можно сделать вывод, что данная нейросеть не различает ассоциативные значения, построенные на метонимическом переносе, и символические значения, мотивационная основа которых представляет собой определенные культурные, исторические и политические фрагменты картины мира различных народов.

При обращении с тем же запросом к чату *Qwen3* (дата обращения 15.07.2025) мы получаем принципиально отличные результаты, т. к. ИИ предлагает не просто семантическую структуру прилагательного *red*, а представляет уже проанализированные данные, которые категоризованы по шести семантическим блокам с указанием мотивационной основы и в некоторых случаях типа дискурса (рис. 2).

🌸 5. Сезонные и эстетические образы

- Рождество и праздники :

Красный — один из главных цветов Рождества (Дед Мороз, рождественские шары, свечи).

- Мода и красота :

Красная помада, ногти, платье — часто символ уверенности, женственности, роскоши.

💜 6. Литературные и метафорические значения

- Смерть, насилие :

В литературе и кино красный часто используется для передачи жестокости, войны или трагедии.

- Жизнь и жизненная энергия :

Цвет крови, сердца, любви — символ жизни и эмоций.

Рис. 2. Семантические блоки 5 и 6 по данным чата *Qwen*

Завершает семантическую справку итоговая таблица со всеми значениями и иллюстративными примерами к ним (рис. 3).

Итоговая таблица:

ЗНАЧЕНИЕ	ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ
Гнев / злость	see red, turn red with anger
Страсть / сексуальность	red dress, red lips
Опасность / тревога	red alert, stop sign, red light
Болезнь / боль	red from pain, flushed face
Смущение	go red in the face
Политика / идеология	red scare, red revolution
Экономика	in the red (в убытке)
Литература / кино	blood, violence, love scenes

Рис. 3. Итоговая таблица значений термина red по данным чата Qwen3

При анализе первого семантического блока «эмоции и страсть» выделяется новое символическое значение «страсть, сексуальность, желание», которое мотивируется ассоциациями с красным цветом одежды или ярким макияжем женщины в мире моды и кино. В качестве иллюстрации предлагается пример: *Shewore a red dress that turned every head*. Значения второго блока «опасность и предупреждение» повторяют значения, которые были представлены в результате запроса в чате ruGPT. Значения третьего блока «физиология и здоровье» не являются символическими, т. к. построены на метонимическом переносе, что было описано выше при анализе данных чата ruGPT.

Значения в четвертом семантическом блоке «культурные и исторические ассоциации» отвечают цели запроса (рис. 4), т. к. нейросеть не только фиксирует данные значения, но и предлагает лингвистические данные, на основе которых они выделены с указанием исторических, политических и других символических ассоциаций, закрепленных в Великобритании и США. Интересно отметить, что символические значения, полученные при аналогичном запросе в чате ruGPT, также построенные на исторических и культурных символах США и КНР чат Qwen3 не генерирует.

4. Культурные и исторические ассоциации

- Английские идиомы и выражения :
 - Caught red-handed* — пойман с поличным.
 - Paint the town red* — весело провести время, устроить гулянку.
 - Red tape* — бюрократия, формальности.
 - In the red* — быть в убытке (финансовый термин).
 - Red flag* — сигнал опасности или проблемы.

- Политика :
 - *Red* иногда ассоциируется с левой идеологией, социализмом или коммунизмом.
Например: *The Red Scare* — антикоммунистическая истерия в США.
- Национальные символы :
 - В британской культуре "*the British redcoats*" — солдаты британской армии.
 - Флаги многих стран содержат красный цвет как символ силы, крови и жертвоприношения.

Рис. 4. Семантический блок 4 по данным чата Qwen

Подводя итоги проводимого лингвистического анализа семантики термина *red* в английском языке с помощью чатов Qwen3 и ruGPT, можно выделить определенные преимущества и недостатки, свойственные обоим ИИ инструментам и каждому в отдельности. Первый вывод, который может сделать каждый исследователь, заключается в том, что использование только одного ИИ инструмента для анализа семантики языковой единицы представляется нецелесообразным, т. к. данные, полученные в результате генерации ответа на один и тот же запрос, отличаются в значительной степени. Более продуктивно использовать данные как минимум двух или даже нескольких ИИ чатов для сбора более объективных данных. По содержательному критерию чат Qwen3 способен генерировать более логичную и развернутую семантическую структуру с указанием мотивационной основы значения и типа дискурса, характерного для его реализации в языке. Также, в отличие от ruGPT, чат Qwen3 использует сами лингвистические данные (идиоматические выражения, иллюстративные примеры), которые позволяют исследователю сделать самостоятельный вывод об объективности генерированной информации. Более того, чат Qwen3 делает попытку предварительного анализа всех значений, группируя их в несколько семантических блоков, что совершенно не свойственно чату ruGPT.

Общими зонами риска для исследователя является неумение чатов различать специфику формирования семантического значения, выделяют только символические значения. Чаты ruGPT и Qwen3 ошибочно выделяют как символические все не прямые значения термина *red*, что является лингвистической ошибкой. Также в качестве недостатка работы с ИИ инструментами на примере анализируемых чатов можно выделить отсутствие ссылок на источники выделяемых значений, что требует организации последующих этапов работы в общем алгоритме проводимого исследования (Мичугина, 2025).

ЛИТЕРАТУРА

1. Мичугина С. В. Денотативное пространство прилагательных цвета в английском языке : специальность 10.02.04 «Германские языки» : автореферат диссертации на соискание ученой степени кандидата филологических наук / Мичугина Светлана Викторовна. – Москва, 2005. – 24 с. – EDN NIMRMR.

2. Мичугина С. В. Антропоцентризм в восприятии цвета / С. В. Мичугина // Когнитивные исследования языка. – 2016. – № 27. – С. 210-216. – DOI 10.20916/2071-9639-2016-27-210-216. – EDN XDDSKH.
3. Мичугина С. В. Терминология цвета в английском языке / С. В. Мичугина // Когнитивные исследования языка. – 2017. – № 30. – С. 753–757. – EDN ZMQJGB.
4. Мичугина С. В. ИИ как инструмент исследования художественного дискурса (на примере чата *gpt*) / С. В. Мичугина // От слова к дискурсу. Материалы межд. науч. конференции. Минск, 15–17 мая 2025. – 2025. – С. 302–305.

УДК 81'33

Николаева Марина Николаевна

кандидат филологических наук,
доцент кафедры английской
филологии
Московский городской
педагогический университет
г. Москва, Россия

Marina Nikolaeva

PhD in Philology,
Associate professor
Moscow City University
Moscow, Russia
nikolaevam@mgpu.ru

Овсянникова Марина Анатольевна

кандидат филологических наук,
доцент кафедры английской
филологии
Московский городской
педагогический университет
г. Москва, Россия

Marina Ovsiannikova

PhD in Philology,
Associate professor
Moscow City University
Moscow, Russia
ovsyannikovama@mgpu.ru

ИНСТРУМЕНТЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ КОРПУСНОГО АНАЛИЗА ИДИОМ С КОЛОРОНИМОМ GREEN: СРАВНИТЕЛЬНЫЙ АСПЕКТ

Данная статья представляет результаты сравнительного анализа применения корпусного анализа на ресурсе english-corpora.org (корпус NOW) и инструментов искусственного интеллекта (Mistral, Deepseek, Perplexity, GigaChat) при отборе лексического материала, в частности, фразеологических единиц с компонентом колоронимом *green* в английском языке. Основное внимание уделено смешанной идиоме *green behind the ears*, которая была выделена с помощью ИИ и верифицирована в корпусе NOW. Авторы анализируют ее происхождение и контексты употребления. Также рассматривается фразеологический бленд *green around the gills with envy*, обнаруженный в корпусе, и его интерпретация с помощью ИИ. Исследование демонстрирует как преимущества, так и ограничения доступных в России ИИ-платформ: недостоверность источников у GigaChat и ограниченность временного охвата данных у Mistral и Deepseek. Однако авторы подчеркивают ценность инструментов ИИ для расширения возможностей лингвистического анализа, в особенности в сочетании с традиционными корпусными методами.

Ключевые слова: искусственный интеллект; корпусный анализ; фразеология; колороним; идиома; бленд.

AI TOOLS FOR CORPUS ANALYSIS OF GREEN-COLOUR IDIOMS: COMPARATIVE STUDY

The article presents the results of a comparative study of corpus data analysis on the resource english-corpora.org (the NOW corpus) and artificial intelligence tools (Mistral, Deepseek, Perplexity, GigaChat) when selecting lexical material, in particular, phraseological units with the colouronym component *green* in English. The main focus is on the mixed idiom

green behind the ears which was identified by AI and verified in the NOW corpus. The authors analyze its origin and usage contexts. The phraseological blend *green around the gills with envy* found in the corpus and its interpretation using AI are also considered. The study demonstrates both the advantages and limitations of the AI platforms available in Russia: the unreliability of sources from GigaChat and the limited time coverage of data from Mistral and Deepseek. However, the authors emphasize AI value for expanding the possibilities of linguistic analysis, especially in combination with traditional corpus methods.

Key words: artificial intelligence; corpus analysis; phraseology; colouronym; idiom; blend.

На территории России доступен агрегатор корпусных текстов English Corpora (<https://www.english-corpora.org>). Наиболее полный корпус на данном сайте – это корпус веб-изданий, который обновляется в режиме реального времени, корпус NOW [1]. Однако с данным ресурсом часто возникают технические проблемы: зависает страница поиска, поиск осуществляется не сразу, часто приходится перезагружать сайт. В настоящий момент с развитием платформ искусственного интеллекта (ИИ) исследователи могут применять новые технологии для лингвистического анализа наряду с традиционными методами [2, 3]. В рамках данного исследования была предпринята попытка воспользоваться доступными в России инструментами искусственного интеллекта для поиска фразеологизмов с колоронимом *green* на английском языке в корпусе NOW.

Колоронимы являются универсальными языковыми единицами, отражающими восприятие окружающего мира человеком, и, более того, влияющими «на изменение восприятия мира и различных человеческих ценностей» [4, с. 15]. Колороним зеленого цвета в современное время привлекает значительное внимание лингвистов. Как отмечает Мишель Пастуро, зеленый цвет, «некогда презируемый, отвергнутый, нелюбимый, сегодня превратился в мессианский цвет» [5, с. 139]. Обладая амбивалентной семантикой, он вызывает как положительные, так и отрицательные ассоциации. К отрицательным чаще всего относят зависть, скупость, ревность, к положительным – свежесть, юность, гармонию, естественность, симпатию и доверие. Интерпретируя смысл колоронимов, «человек опирается на прошлый опыт, составляющий часть индивидуального и коллективного знания» [6, с. 395], представленного, в частности, в фразеологических выражениях, или идиомах.

Как уже было сказано выше, корпусный менеджер English Corpora не всегда работает корректно, поэтому поиск совокупности фразеологизмов по ключевому слову *green* проблематичен. В связи с этим на первом этапе для выделения идиом с колоронимом *green* была использована платформа Perplexity. Платформа предложила 16 идиом, из которых внимание авторов статьи привлекла идиома *green behind the ears*.

На следующем этапе сама идиома была верифицирована через корпус NOW и найдено 47 случаев ее употребления с 2010 года (начало работы корпуса) по 2025 год. Приведем некоторые примеры:

*He could see we were a bit **green behind the ears** so he realised we needed help* (2013).

*I was fairly new in the job, and about as **green behind the ears** as it gets* (2018).

*I was a young researcher back then, still **green behind the ears*** (2025).

Данная идиома представляет собой смешанную идиому (idiom blend), которая могла возникнуть либо как слияние семантически близких *greenhorn* и *be wet behind the ears*, либо оказалась в английском языке как калька от немецкого фразеологизма *grün hinter den Ohren*. Знание схожих фразеологических оборотов и контексты, найденные в корпусе, делает значение очевидным: молодой, неопытный, желторотый, еще зеленый.

Сайт English-Corpora.org обещает внедрить ИИ в работу сайта уже в августе 2025 года [7], однако на момент написания статьи функции анализа коллокаций с помощью нейросетей не было. Авторы статьи приняли решение проверить, могут ли доступные в России нейросети (Mistral AI, Deepseek AI, Perplexity и GigaChat) найти контексты использования определенного выражения за конкретный временной период и провести дальнейший анализ. Сначала был задан следующий промпт: *найди контексты использования идиомы green behind the ears за 2025 год*. Так, удалось выяснить, что у моделей ограничена база знаний: у Mistral – до 1 ноября 2024 года, у Deepseek – до июля 2024 года: *Поиск примеров именно за текущий год требует анализа огромного массива данных, генерируемых ежесекундно, что недоступно в моей текущей модели* [8, 9]. Запрос был изменен на 2023 год, который входил во временной период базы данных, указанный моделями, но контексты употребления получить не удалось.

GigaChat дал на первый взгляд полный ответ с источниками, однако при проверке 10 источников ни один не был настоящим: одни страницы не открывались, на других отсутствовал цитируемый текст [10].

Необходимо отметить положительную сторону использования нейросетей в данном исследовании. Все ИИ платформы дали отсылку к самому известному случаю публичного использования этой идиомы: речь Барака Обамы в 2008 году, когда он сказал, что его оппонент Джон Маккейн считает Обаму молодым и неопытным политиком: *Now, Senator McCain suggests that somehow, you know, I'm **green behind the ears** and, you know, I'm just spouting off, and he's somber and responsible* [11].

Благодаря проверке частотности лексических единиц с помощью Google Books Ngram Viewer, удалось выяснить, что идиома *green behind the ears* приведена в книге «The Synonym Finder» как синоним для *unaccustomed* в значении *not used to, not ready for, not up to; uninitiated, unpracticed, inexperienced; green, raw, wet behind the ears* уже в 1978 году [12, p. 1271].

Интересно, что корпусный поиск по отдельным идиомам с колоронимом *green* наряду с множественными употреблениями идиом *green around the gills*

и *green with envy* предложил единичное употребление окказионального фразеологического бленда: *green around the gills envy*. В статье жанра *Agony Aunt* американского онлайн-издания *Slate Magazine – Politics, Business, Technology and the Arts* 2025 года приводится отрывок из письма, адресованного лицу, представленного псевдонимом *Prudence* (рус. благоразумие, осторожность, расчетливость): *What do you do with **green around the gills envy**? My partner inherited a five-bedroom house, and we plan to renovate it with grand dreams* [13]. При проверке данной идиомы в *Google Books Ngram Viewer* результаты получены не были.

На запрос с анализом данной идиомы нейросети дали очевидный ответ: идиомы *green around the gills* и *green with envy* имеют разные значения и контексты употребления. Их объединяет только слово *green*, но семантически они не связаны [8, 9, 10]. Однако, интересно отметить, что *Perplexity*, дав подробный разбор идиом, из которых состоит бленд, указал на возможность использования такого фразеологизма: *Совместное употребление (редко, но возможно). Иногда можно встретить оба выражения в одном тексте, когда описывают человека, который выглядит плохо из-за зависти или сильных эмоций: He was **green around the gills with envy** when he saw his rival's promotion (Он выглядел болезненно и бледно от зависти, когда увидел повышение своего соперника)* [14]. На дополнительный запрос о конкретных примерах в литературе или фильмах *Perplexity* дал еще несколько примеров, однако указал, что они «стилизированные», объяснив это так: *Реальных цитат с одновременным употреблением идиомы **green around the gills** и слова **envy** в классической литературе или известных произведениях найти очень сложно, так как эти выражения обычно используются отдельно и в разных контекстах: **Green around the gills** – идиома, связанная с физическим состоянием (тошнота, бледность); **Envy** – эмоциональное состояние, чаще описываемое через метафоры (**green-eyed monster**) или отдельные выражения* [14]. В указанном выше примере из американского онлайн-издания происходит обыгрывание значения колоронима *green*: колороним *green* в составе идиомы *green around the gills* служит стилистическим фразовым эпитетом, находящимся в акцентированной препозиции к определяемому слову *envy*, и выполняет функцию передачи эмоционального состояния сильной зависти, приводящей человека к болезненному недомоганию. Интерпретация смысла употребления идиомы с колоронимом в таком ключе согласуется с восприятием зависти в западноевропейской и американской культуре как ассоциируемой с зеленым цветом, в отличие от русской культуры, в которой зависть может характеризоваться как белая или черная.

Таким образом, следует отметить, что применение корпусного анализа и инструментов ИИ, взаимодополняющих друг друга при рассмотрении фразеологизмов, в нашем случае с колоронимом *green*, значительно расширяет экспериментальные возможности и представляется перспективным в дальнейших лингвистических исследованиях.

ЛИТЕРАТУРА

1. NOW Corpus (News on the Web) : официальный сайт. URL: <https://www.english-corpora.org/nw/> (дата обращения: 10.07.2025).
2. Овсянникова М. А., Николаева М. Н. Нейросети как инструмент анализа текста: сравнительный аспект // Язык в эпоху цифровых трансформаций и развития искусственного интеллекта : сборник научных статей по итогам международной научной конференции, Минск, 23–24 октября 2024 года. Минск : Минский государственный лингвистический университет, 2024. С. 570–575.
3. Петрова И. М., Иванова А. М., Никитина В. В. Современные цифровые технологии в лингвистических исследованиях : Учебное пособие для обучающихся по направлению «Лингвистика»; Московский городской педагогический университет, Институт иностранных языков. Москва : Языки Народов Мира, 2022. 259 с.
4. Сластикова Т. В., Черкашина Е. И. Цвет и цветообозначение в лингвистических исследованиях. Москва : Языки народов мира, 2021. 240 с.
5. Пастуро М. Зеленый. История цвета; пер. с фр. Н. Кулиш. 4 изд. Москва : Новое литературное обозрение, 2024. 168 с.
6. Николаева М. Н. Концептуализация колоронимов синего цвета в итальянском языке // Когнитивные исследования языка. 2025. № 1–1(62). С. 394–398.
7. English Corpora : официальный сайт. URL: <https://www.english-corpora.org> (дата обращения: 10.07.2025).
8. Deepseek : официальный сайт. URL: <https://chat.deepseek.com> (дата обращения: 10.07.2025).
9. Mistral AI : официальный сайт. URL: <https://chat.mistral.ai> (дата обращения: 10.07.2025).
10. GigaChat : официальный сайт. URL: <https://giga.chat> (дата обращения: 10.07.2025).
11. The Second Presidential Debate: Obama vs. McCain. URL: <https://archive.nytimes.com/www.nytimes.com/elections/2008/president/debates/transcripts/second-presidential-debate.html> (дата обращения: 10.07.2025).
12. Urdang, Laurence., LaRoche, Nancy. The Synonym Finder. Rodale Books, 1978. 1361 p.
13. Help! My Fiancé Took a DNA Test. URL: <https://slate.com/advice/2025/06/family-advice-fiance-wedding-dna-test.html> (дата обращения: 10.07.2025).
14. Perplexity : официальный сайт. URL: <https://www.perplexity.ai> (дата обращения: 10.07.2025).

УДК 81'33

Прокофьева Татьяна Владимировна

РГПУ им. А. И. Герцена
г. Санкт-Петербург, Россия

Tatyana Prokofyeva

Herzen University
Saint-Petersburg, Russia
tatyana.prokofyeva.herzen@mail.ru

Белоглазова Елена Владимировна

доктор филологических наук, доцент
РГПУ им. А. И. Герцена
г. Санкт-Петербург, Россия

Elena Beloglazova

PhD in Philology, professor
Herzen University
Saint-Petersburg, Russia
evbeloglazova@herzen.spb.ru

АНАЛИЗ ТОНАЛЬНОСТИ РУССОЦЕНТРИЧЕСКОГО ДИСКУРСА

Статья посвящена исследованию тональности руссоцентрического дискурса в современном медиапространстве. Актуальность работы определяется усилением интереса к национальной идентичности и идеологическим конструктам России, транслируемым через язык. Целью исследования является анализ тональности иноязычных текстов, посвященных России, для выявления ключевых тем, стереотипов и эмоциональных оценок. В качестве методологической основы применяются методы компьютерной лингвистики, в частности автоматический анализ тональности текста с использованием модель VADER/NLTK и моделью RoBERTa. Эмпирической базой для исследования послужил ad-hoc корпус заголовков книг о России (новостных и публицистических статей, блогов и социальных сетей, названий книг). Результаты анализа тональности разделены на «позитивные», «нейтральные» и «отрицательные». Также рассматриваются существующие ограничения используемых инструментов для автоматического анализа тональности. По данным исследования были выявлены закономерности работы используемых моделей и их особенности, которые необходимо учитывать при использовании данных инструментов.

Ключевые слова: оценка; оценочное суждение; тональность; ценность; анализ тональности; руссо-центрический дискурс.

SENTIMENT ANALYSIS OF RUSSIA-CENTRED DISCOURSE

The article is devoted to the study of sentiment in Russia-centric discourse in the contemporary media landscape. The relevance of the research is determined by the growing interest in national identity and Russia's ideological constructs transmitted through language. The aim of the study is to analyze the sentiment of foreign-language texts about Russia in order to identify key themes, stereotypes, and emotional evaluations. As a methodological basis, methods of computational linguistics are applied, in particular automatic sentiment analysis of text using the VADER/NLTK model and the RoBERTa model. The empirical basis for the study was an ad-hoc corpus of book titles about Russia (including news and journalistic articles, blogs, and social media, as well as book titles). The sentiment analysis results are categorized into "positive", "neutral", and "negative". Existing limitations of the tools used for automatic

sentiment analysis are also considered. The study identified patterns in the performance of the models used and their specific characteristics, which must be taken into account when employing these tools.

Key words: evaluation; evaluative judgment; sentiment; value; sentiment analysis; Russia-centred discourse.

В современном мире в эпоху постглобализации продолжается глобализация информационного пространства, что, с учетом пристального внимания международного сообщества к России обуславливает актуальность исследования тональности руссоцентрического дискурса. Иноязычные описания России напрямую формируют общественное мнение как внутри страны, так и за её пределами, при издании книг и публикации новостных репортажей в других государствах. В условиях информационных войн и культурных различий формальный анализ оценочности иноязычных описаний России становится важным инструментом для понимания внешнего восприятия образа страны и её места в мире.

“Тайна оценочного значения на протяжении веков волновала философов, логиков и лингвистов”, констатирует Н. Д. Арутюнова, оценивая проблему как “твердый орешек”. С момента этой констатации наука, в том числе лингвистика, испытала существенные перемены, связанные с появлением технологий автоматизированного анализа текста с привлечением больших массивов данных.

Цель данной работы заключается в оценке адекватности автоматизированного анализа тональности на самобытном материале текстов иноязычного руссоцентрического дискурса.

Теоретический фундамент

Исследование имеет две теоретико-методологические предпосылки.

Во-первых, анализ тональности, являясь прикладной задачей, направленной на извлечение из текста оценочного компонента (“*opinion mining*”), требует уяснения лингвистической природы оценки.

Н. Н. Болдырев выделяет оценку как одну из трех типов категорий в языке. Под оценочной категоризацией исследователь понимает «группировку объектов и явлений по характеру их оценки в соответствующие оценочные классы и категории, то есть систему оценочных категорий (статический аспект), или мыслительное соотнесение объекта или явления с определенной оценочной категорией (динамический аспект)» [Болдырев 2002: 104]. Именно включение в поле оценочности динамического аспекта и делает оценку столь интересной проблемой для изучения: оценка есть не только неотъемлемое свойство определенного класса языковых единиц, но может возникать в коммуникации как косвенный прагматический эффект высказывания.

Из сказанного явствует, что оценка проявляется на всех уровнях речевой коммуникации, она может выражаться через различные лексические единицы, грамматические и стилистические средства, которые позволяют

выражать субъективное отношение субъекта к объекту оценки. Согласно исследованиям Е. М. Вольф, оценочные суждения в языке включают в себя субъективный (отражает личное отношение говорящего к предмету) и объективный (отражает признаки самого объекта оценки) [Вольф, 2002].

Словами, специализирующимися на выражении оценки, принято считать прилагательные, оценочные существительные и глаголы с явной оценкой, которые выражают положительное или отрицательное оценочное значение, данные слова служат основными показателями оценки в языке. Помимо данных слов, выделяют такие языковые явления, которые содержат в себе оттенки оценки:

1. Уменьшительно-ласкательные формы слов («глаза» – «глазоньки»);
2. Суффиксальные формы слов («дом» – «домик»);
3. Коннотация («политик» – «политикан»);
4. Различные интенсификаторы («очень», «довольно», «ультра-» и т. д.);
5. Междометия («Браво!», «Фу!» и т. д.);
6. Модальность;
7. Экспрессивные средства языка (интонация, грамматические конструкции, фразеологизм, метафора, сравнение и т. д.) [Вольф, 2002; Темигразина, 2015].

Оценка и тональность тесно связаны друг с другом, но оценка имеет более широкий спектр смысловых и прагматических аспектов (когнитивные, эмоциональные и поведенческие), в то время как тональность – более узкое понятие, является выражением эмоциональной окраски высказывания, его субъективной модальности и направленности (положительной, нейтральной, отрицательной) [Чернявская, 2020; Нефедов, 2019]. Таким образом, можно считать, что тональность есть частный случай оценки. Тональность формируется на основе оценочных суждений, но не всякая оценка выражается в тональности. Она зависит от оценки и отражает ее эмоциональный аспект. Анализ взаимосвязи оценки и тональности позволяет утверждать, что оценочные структуры лежат в основе формирования тональности текста, а их интерпретация требует комплексного подхода, учитывающего как лингвистический, так и культурно-прагматический контекст.

Во-вторых, исследование тональности представляет интерес с точки зрения изучения своеобразия иноязычного руссоцентрического дискурса, который можно определить как «особое использование языка для вербализации представлений о России и русской культуры, складывающееся в определенном дискурсивном сообществе под влиянием факторов культурной и идеологической дистанции». [Белоглазова, 2025: с 134] Определенные особенности данного дискурса описаны в работах Е. В. Белоглазовой и Н. А. Осьмак, где авторы отмечают своеобразные для дискурса оценочно-заряженные концептуальные доминанты.

Материалы и методы

Основными инструментами для автоматического анализа тональности стали модель VADER/NLTK и модель RoBERTa.

Для работы с моделью VADER/NLTK были заданы определенные параметры, относительно которых она определяла какой тип тональности необходимо выставить для каждого текста. Так, текстопределялся:

1. «положительным», если общее значение текста заголовка равнялось или превышало показатель 0,5;
2. «нейтральным», если общее значение текста заголовка находилось в диапазоне от -0,5 до 0,5;
3. «отрицательным», если общее значение текста заголовка равнялось или было ниже -0,5.

При использовании стандартных настроек модель RoBERTa не выделяет тексты с нейтральной тональностью, разделяя их только на положительные и отрицательные. Однако, можно самостоятельно задать параметры, в рамках которых тексты будут определены как тексты с нейтральной тональностью (в данном случае значение $< 0,6$).

Для проведения анализа тональности был выбран ad-hoc корпус заголовков англоязычных статей и книг, посвященных России, Rossica-T, включающий 662 заголовка. Выбор материала обусловлен минимальной значимостью в нем контекста, что позволяет в наиболее чистом виде оценить действие словарного подхода к анализу тональности.

Результаты исследования

Результаты проведения анализа тональности представлены в виде круговых диаграмм с процентным соотношением автоматически определенных положительной, нейтральной и отрицательной тональностей (рис. 1 и рис. 2).

Результаты анализа тональности (VADER/NLTK)

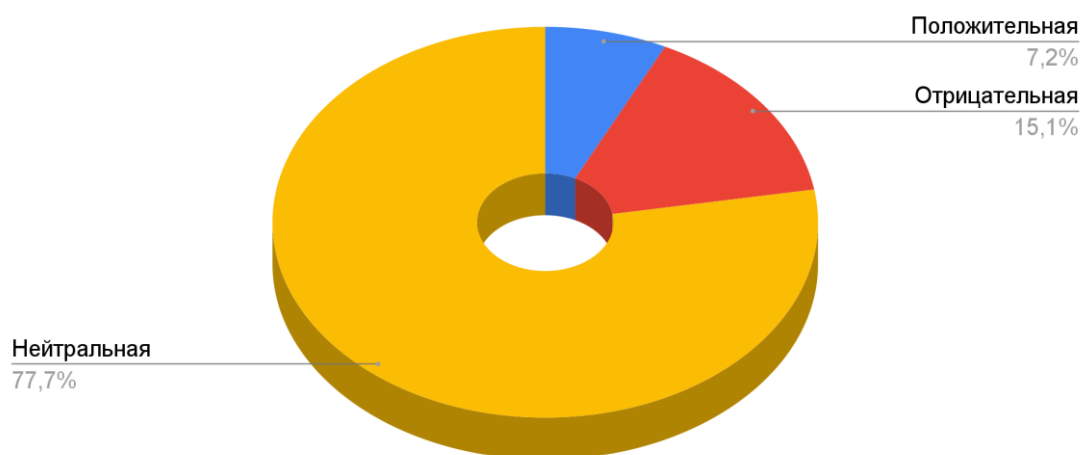


Рис. 1. Диаграмма с результатами анализа тональности (VADER/NLTK)

Результаты анализа тональности (модель RoBERTa)

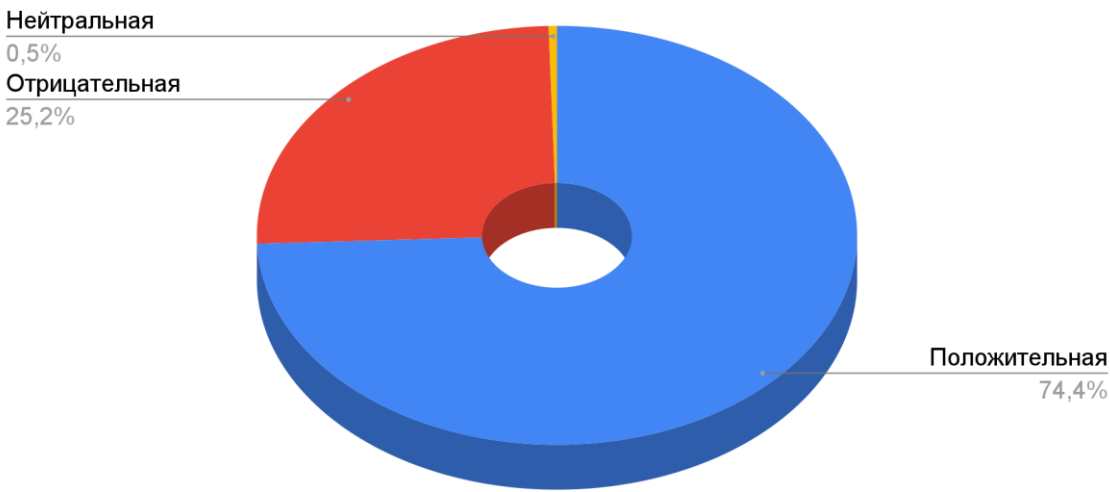


Рис. 2. Диаграмма с результатами анализа тональности (модель RoBERTa)

При сравнении результатов анализа тональности выбранными инструментами можно заметить большую разницу в соотношении текстов. Так, модель RoBERTa выдает более точные результаты, где тексты положительной тональности занимают 74,4 %, с нейтральной тональностью составляют 0,5 %, а с отрицательной – 25,2 %. В результате анализа с помощью данной модели можно увидеть, что положительная тональность превалирует над отрицательной. В то же время модель VADER/NLTK выделяет большую часть заголовков, как тексты с нейтральной тональностью – 77,7 %, соотношение текстов с положительной и отрицательной тональностями противоположно тому, что показывает диаграмма модели RoBERTa, заголовки с отрицательной тональностью составляют 15,1 %, а с положительной – 7,2 %.

Данные, полученные в результате исследования, подтверждают актуальность применения методов автоматического анализа тональности для исследования образа России в англоязычном дискурсе. Использование двух инструментов – библиотеки NLTK и модели RoBERTa – позволило выявить как общие закономерности, так и значительные расхождения в оценке одного и того же корпуса текстов.

С одной стороны, модель RoBERTa показала большую чувствительность к эмоционально окрашенной лексике, что было отражено в процентном соотношении текстов с положительной тональностью (74,4 %) и отрицательной тональностью (25,5 %), в то время как заголовки с нейтральной тональностью составили всего 0,5 %. Это может быть связано как с особенностями обучающего корпуса модели, так и с контекстной глубиной анализа, обеспечиваемой архитектурой данной модели.

С другой стороны, модель VADER/NLTK показала большой процент текстов с нейтральной тональностью – 77,7 %, это может быть обусловлено тем, что данная модель ориентирована на словарные значения и ограничена в способности считывать дискурсивный контекст.

При сравнении моделей можно сделать вывод, что для более точного анализа больше подходит модель RoBERTa, так как полученные результаты обладают большей точностью, но для анализа данных требуется больше времени. Модель VADER/NLTK же подходит для первичного анализа, либо для определенных исследований, в которых необходимо задавать собственные параметры определения тональности при анализе. RoBERTa выдает результаты с выраженной бинарной полярностью, в то время как VADER демонстрирует склонность к оценочной умеренности.

При необходимости анализа большого количества данных, автоматический анализ тональности значительно сокращает время, необходимое для его реализации. Использование автоматического анализа тональности позволит эффективнее изучить иноязычного руссоцентрического дискурс. Однако, необходимо учитывать особенности используемых инструментов, настройку моделей с учетом специфики культурного, жанрового и идеологического контекста, а также ведение контроля качества полученных данных.

ЛИТЕРАТУРА

1. Арутюнова Н. Д. Типы языковых значений: Оценка. Событие. Факт. / Арутюнова Н. Д., Степанов Г. В. – М.: Наука, 1988. – 339 с.
2. Болдырев Н. Н. Структура и принципы формирования оценочных категорий // С любовью к языку: Сб. Научных трудов. – Москва-Воронеж: ИЯ РАН, Воронежский государственный университет, 2002. – С. 103–114.
3. Вольф Е. М. Функциональная семантика оценки / Е. М. Вольф – 2-е изд., доп. – М.: Едиториал УРСС, 2002. – 280 с. – Текст: электронный – URL: <https://reallib.org/reader?file=579811&pg=111> (дата обращения: 12.01.2025).
4. Темиргазина З. К. Лингвистическая аксиология: оценочные высказывания в русском языке: монография / З. К. Темиргазина. – 2-е изд., стер. – Москва: ФЛИНТА, 2015. – 247 с. – ISBN 978-5-9765-2118-6. – Текст: электронный // Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/70403> (дата обращения: 04.01.2025). – Режим доступа: для авториз. Пользователей.
5. Чернявская В. Е. Язык оценок в научном дискурсе: терминологическое поле и методологические подходы / В. Е. Чернявская // Вестник Пермского университета. Российская и зарубежная филология. – 2022. – Т. 14, № 3. – С. 44-55. – DOI 10.17072/2073-6681-2022-3-44-55. – EDN SVXEVH.
6. Нефедов С. Т. Варьирование оценки в коммуникативных практиках научного дискурса / С. Т. Нефедов // Вестник Санкт-Петербургского университета. Язык и литература. – 2021. – Т. 18, № 4. – С. 760–778. – DOI 10.21638/spbu09.2021.408. – EDN HPNSWE.

7. Белоглазова Е. В. Образ России в зеркале японоязычных кинокритических рецензий / Белоглазова Е. В., Сосункевич В. О. // Социальные и гуманитарные науки на Дальнем Востоке. – 2025. – Том 22, N 1. – С. 132–137. –DOI: 10.31079/1992-2868-2025-22-1-132-137.
8. Белоглазова Е. В. Rossica и Belarossica: общие и культурно-специфические маркеры дискурсов иноязычного описания России и Беларуси / Белоглазова Е. В., Генидзе Н. К. ; RossicaV. Belarossica: Universal and Culture-Specific Features of Russia- and Belarus-Centered Discourses // Вопросы когнитивной лингвистики. – 2023. –N 4. – С. 108–115.
9. Белоглазова Е. В., Осьмак Н. А., Шувалова Е. К. Формальные маркеры и содержательные доминанты руссоцентрического дискурса: кросс-языковое корпусное исследование // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2024. Т. 22, № 1. С. 105–119. DOI 10.25205/1818-7935-2024-22-1-105-119.
10. Осьмак Н. А. Образ Сибири в современном финноязычном новостном дискурсе = The Image of Siberia in Contemporary Finnish-language News Discourse / Осьмак Н. А. // Журнал Сибирского федерального университета. Гуманитарные науки. – 2025. – Том 18, N 1. – С. 21–33. – URL: <https://elibrary.ru/item.asp?id=80336744>.
11. Семина Т. А. Тональность текста: синтаксические паттерны выражения отношений между сущностями : специальность 10.02.19 «Теория языка» : защищена 20.05.2020 / Семина Т. А.; Российский университет дружбы народов; научный руководитель Максименко О. И. – Мытищи, 2020. – 24 с. – Текст: электронный // Elibrary: электронно-библиотечная система. – URL: https://www.elibrary.ru/download/elibrary_45677042_93462808.pdf (дата обращения 15.04.2025). – Режим доступа: для авториз. Пользователей.
12. More than a Feeling: Accuracy and Application of Sentiment Analysis / Jochen Hartmann, Mark Heitmann, Christian Siebert, Christina Schamp // International Journal of Research in Marketing. – 2023. – Vol. 40, Issue 1. – pages 75–87. – ISSN 0167-8116.

УДК 81'324:004.8:32

Рябченко Наталья Анатольевна

кандидат политических наук,
доцент кафедры социальной работы
психологии и педагогики высшего
образования; руководитель «Меж-
дисциплинарной научно-исследова-
тельской лаборатории прикладной
лингвистики и сетевых исследований
публичной политики»
Кубанский государственный
университет
г. Краснодар, Россия

Natalia Ryabchenko

PhD in Political Science,
Associate Professor, Department of
Social Work, Psychology, and Higher
Education Pedagogy;
Head, Interdisciplinary Research
Laboratory of Applied Linguistics and
Network Studies in Public Policy
Kuban State University
Krasnodar, Russia
rrrnatali@mai.rul

Малышева Ольга Петровна

кандидат филологических наук,
доцент кафедры английской
филологии
Кубанский государственный
университет
г. Краснодар, Россия

Olga Malysheva

PhD in Philology,
Associate Professor, Department
of English Philology
Kuban State University
Krasnodar, Russia
malysheva_op@mail.ru

МЕТОДОЛОГИЧЕСКИЕ СДВИГИ В ПРИКЛАДНОЙ ЛИНГВИСТИКЕ ПОД ВЛИЯНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В условиях стремительной цифровизации и развития искусственного интеллекта (ИИ) прикладная лингвистика претерпевает принципиальные изменения. Переход от обработки естественного языка к генеративным системам, синтезирующим контент, требует пересмотра теоретико-методологических рамок. Дискурс-анализ, акторно-сетевая теория, когнитивно-лингвистические и этико-правовые подходы объединяются в единый каркас для анализа текстовых и мультимодальных данных. Статья предлагает обзор существующей методологии и перспективы прикладных лингвистических исследований, предполагающих междисциплинарный подход и использование корпусно-сетевых методов, графовых нейросетей и моделирования предикационных сетей. В качестве показателя качественных изменений исследуется кейс «Парламентские выборы в Индии 2024 г.», ставший прецедентом дискурсивного сдвига, определившего начало эволюции жанра политического дискурса (в том числе в связи с массовым распространением дипфейков) и изменившего представления о прагматике политической коммуникации. Основными выводами являются: 1) ИИ выступает инструментом и актором в политическом дискурсе; 2) совместное продуцирование человеком и машиной трансформирует категории «источник», «аудитория» и «интенция»; 3) комплексное сочетание дискурс-аналитических, сетевых и экспериментальных методов позволяет адекватно фиксировать и прогнозировать новые формы политической коммуникации в эпоху генеративных технологий.

Ключевые слова: корпусная лингвистика; искусственный интеллект; политический дискурс; сетевой анализ; графовые нейросети; генеративные языковые модели.

AI-DRIVEN METHODOLOGICAL SHIFTS IN APPLIED LINGUISTICS

Rapid digitalization and the advancement of artificial intelligence (AI) allowed for fundamental transformations of applied linguistics. The shift from traditional natural language processing to generative systems capable of synthesizing content determines a thorough reconsideration of existing theoretical and methodological frameworks. Discourse analysis, actor-network theory, cognitive linguistic approaches, and ethical-legal perspectives are being integrated into an integral framework for the examination of textual and multimodal data. This article offers a review of current methodologies and outlines future directions for applied linguistic research that adopt an interdisciplinary stance and employ corpus-based network methods, graph neural networks, and predication-network modeling. As an indicator of qualitative change, the case of the 2024 Indian parliamentary elections is examined, which has become a precedent for a discursive shift marking the evolution of the political discourse genre – regarding the widespread dissemination of deepfakes – and has altered the conception of the pragmatics of political communication. Key findings are as follows: AI functions both as a tool and as an active agent in political discourse; the collaborative co-production by humans and machines transforms the categories of “source,” “audience,” and “intent”; the application of discourse-analytic, network, and experimental methods enables effective documentation and forecasting of emerging forms of political communication in the era of generative technologies.

Key words: corpus linguistics; artificial intelligence; political discourse; network analysis; graph neural networks; generative language models.

В эпоху цифровых трансформаций и стремительного развития технологий искусственного интеллекта (ИИ) прикладная лингвистика переживает качественный сдвиг. В контексте гуманитарных исследований под ИИ следует понимать комплекс алгоритмов и моделей машинного обучения, предназначенных для автоматизации процессов сбора, предобработки, семантической разметки и последующего анализа лингвистических данных с целью разработки систем, обеспечивающих текущую и предиктивную аналитику в приоритетных сферах стратегического развития государства и общества: электоральных кампаниях (моделирование предпочтений избирателей), маркетинговых исследованиях (анализ тональности потребительских откликов) и мониторинге общественно-политических процессов (прогнозирование протестной активности и оценки социальной напряжённости) как в онлайн-, так и в офлайн-пространстве.

Исследовательский фокус сдвигается с чисто инструментального (например, аналитики больших корпусов текстов для выявления скрытых прагматических паттернов), на выработку готовых ИИ решений для власти и бизнеса – разработку генеративных стратегий производства смыслов для управляемого установления повестки дня или моделирования реакции электоральных групп. При этом парадигма лингвистических исследований охватывает уже не линейные модели влияния, но сетевые, мультиагентные и гибридные экосистемы дискурса.

Наряду с методологической трансформацией, современные тенденции демонстрируют, что ИИ становится не просто средством анализа дискурса,

но полноправным актором, определяющим социальные практики и трансформирующим пространство публичной политики – генеративные системы ИИ выступают сценаристами, спичрайтерами, медиаторами и распределителями внимания, формируя новые жанры политического высказывания, модифицируя структуру аргументации и переопределяя границы авторства. Для прикладной лингвистики подобные тенденции означают появление нового эмпирического материала, необходимость переоценки базовых категорий «источник», «аудитория», «интенция», «интерпретация», а также разработки новой теоретической рамки исследования политического и медиадискурса.

Наглядным примером начала лингвистического поворота, спровоцированного развитием технологий ИИ, служат парламентские выборы в Индии 2024 года. Широкое применение генеративных моделей – от синтетических роликов с «воскрешенными» политиками и клонированными голосами до таргетированных чат-ботов и автоматизированных кампаний в мессенджерах [1] – показало перспективы развития лингвистики и породило ряд исследовательских вопросов: каковы лингвистические маркеры «синтетической» политической речи; как генеративные модели влияют на риторику и прагматику обращения к различным электоральным сегментам; каковы сетевые свойства распространения ИИ-сгенерированного контента по сравнению с «органическим»; каким образом меняется доверие к источнику и механизмы легитимации высказывания в условиях машинного соавторства.

Цель данной статьи – систематизировать основные направления исследований ИИ в политической лингвистике, предложить теоретико-методологическую рамку для анализа лингвистических данных и описания современных социально-политических и социально-культурных процессов.

Обзор литературы позволяет заключить, что теоретико-методологическая рамка исследований политического дискурса с применением ИИ формируется несколькими взаимодополняющими подходами: дискурс-анализ, сетевая теория, этические и когнитивно-лингвистические подходы, которые задают вопросы о природе идеологических фреймов, механизмах диффузии языковых паттернов и ответственности за создаваемый контент [2]. На их основе прикладные методы реализуются через корпусные NLP-пайплайны с трансформерами [3], графовые нейросети для моделирования сетевой диффузии «органического» и ИИ-генерированного дискурса [4], боты [5], сравнительный анализ многоязычных LLM [6], а также мультимодальные модели CLIP для объединения вербальных и невербальных компонентов политических сообщений [7]. Экспериментальные протоколы с генеративными языковыми моделями создают «what-if» сценарии и симуляции реакции различных электоральных сегментов [8], в то время как детекция дипфейков и криптографические подписи интегрируются в инфраструктуру доверия для противодействия дезинформации [9]. Стандарты и строгие протоколы валидации (Explainable AI) обеспечивают воспроизводимость и интерпретируе-

мость результатов, связывая эмпирические находки с концептуальными гипотезами [10]. Таким образом, искусственный интеллект сегодня не просто расширяет аналитический арсенал политической лингвистики, но и трансформирует сам объект дисциплины – политический дискурс.

Прошедшие парламентские выборы в Индии – 2024 года стали точкой формирования основных концептуальных направлений развития исследований в области политической лингвистики и искусственного интеллекта, которыми являются:

- 1) Многоязычный дискурсивный сдвиг.
- 2) Жанровая эволюция политического дискурса.
- 3) Прагматический потенциал персонализации общения.
- 4) Создание корпусов «дипфейк – оригинал».

«Многоязычный дискурсивный сдвиг» как концептуальное направление в прикладной лингвистике предполагает ассимиляцию взглядов и идей ввиду когерентности семантических и прагматических структур различных языков, что обеспечивает единое восприятие политических сообщений в многоязычном сообществе. Подобная практика возможна при условии одновременного создания и распространения обращений на множестве языков и диалектов. Многоязычный дискурсивный сдвиг способствует формированию целостного нарратива, в котором адаптация политических посланий к особенностям локальных языков сохраняет их основную идею и прагматический эффект.

Подобный сдвиг был обусловлен первым в истории прецедентом перевода политических обращений к гражданам одновременно на все языки и диалекты многоязычной Индии. Результатом дискурсивного сдвига стала невероятно высокий уровень политического участия и явка на избирательных участках. Реализация подобного проекта стала возможна благодаря привлечению нативной языковой платформы Bhashini [11], обеспечивающей синхронный перевод между всеми основными индийскими языками, английским и хинди. С ее помощью был сформирован обширный корпус параллельных политических сообщений (речей, слоганов, агитационных роликов), проходящих через автоматический перевод. Подобный прецедент создает уникальные условия для изучения широкого круга исследовательских вопросов, охватывающих вопросы модальности и эпистемической маркированности лексики (изменение маркеров уверенности, сомнения или долженствования при машинном переносе); эволюции метафорики и фреймов (перенос культурно специфичных метафор, их «обеднение» или «усиление» в целевом языке); семантической деривации и прагматических потерь (анализ импликаций, пресуппозиций или юмора). Методами анализа эмпирических данных в подобном разрезе становятся сопоставительный корпусный анализ, автоматическая детекция метафор, вычислительная прагматика (анализ модальных операторов), эмбединги предложений для измерения смещения смыслов, ручная экспертная аннотация выборок для калибровки моделей. При этом

приоритетными исследовательскими вопросами становятся возможность формализации культурно-специфических маркеров, оценка качества перевода и выявление корреляции между качеством перевода и эффективностью политического послания в многоязычном электоральном пространстве.

Жанровая эволюция политического дискурса обусловлена тем, что дипфейк-послания и синтетические ролики формируют новую синтетическую форму – гибридные тексты, в которых вербальный компонент (скрипт), вокальный (клонированный голос) и визуальный (сгенерированное видео) функционируют как единое целое и воспринимаются аудиторией нераздельно. Такая интеграция каналов порождает ряд теоретических вызовов. Во-первых, требуется переосмысление понятий иллюкутивной силы и перлокутивного эффекта – под вопросом авторство самой интенции политического послания (политик, команда разработчиков или сама модель). Параллельно происходит реконфигурация жанровых признаков: сатирический ролик, «псевдо-интервью» или мем-видео в условиях машинной генерации меняют жанровые ожидания аудитории, заставляя пересматривать критерии жанровой классификации. Особого внимания требует изучение комического эффекта и способов формализации маркеров иронии для автоматического анализа. Методологическими решениями могут стать мультимодальная разметка (текст, аудио, видео), применение CLIP-подобных моделей, риторический и жанровый анализ, а также эксперименты с восприятием (опросы, ай-трекинг) для измерения перлокутивного эффекта. Открытыми исследовательскими вопросами остаются структура аргументации в гибридных сообщениях, возможность описания прототипического дипфейк-жанра, а также отношение аудитории к «игре с реальностью» в условиях постправды.

Прагматический потенциал персонализированного контента и политической коммуникации является одним из приоритетных направлений, поскольку эмпирически фиксируется феномен индивидуализированных обращений, которые генерируются автоматически и массово распространяются через мессенджеры и чат-боты [12]. В прагматическом фокусе оказывается адресный фрейминг: выбор аргументов, тональности и ценностных ориентиров варьируется в зависимости от социально-демографических характеристик адресата. В связи с этим возникают этические вопросы: где проходит граница между допустимой адаптацией и манипуляцией, нарушающей автономию адресата? Методологическим решением может стать выявление персонализированных шаблонов (promptpatterns), кластеризация сообщений по параметрам адресата, А/В-тесты влияния на поведенческие индикаторы (клики, репосты, намерение голосовать) и каузальный анализ для оценки эффекта персонализации. Лингвопрагматический аспект политической коммуникации является не менее перспективным направлением исследования, поскольку позволит выявлять механизмы усиления убеждающего воздействия, как и моделировать обратные эффекты, такие как отторжение или снижение доверия, возникающие, например, при чрезмерной персонализации.

Создание корпусов «дипфейк – оригинал» является важной, но амбициозной задачей, поскольку для обучения и верификации прозрачных моделей автоматического определения позиции, выявления пропаганды и распознавания синтетического контента необходимы стандартизированные эталонные корпуса, содержащие пары «оригинал – дипфейк». Подход к построению корпуса предполагает несколько этапов. Сначала осуществляются сбор и верификация данных: идентификация исходного материала, проверка подлинности посредством фактчекинга и обратного поиска, фиксация метаданных о времени и каналах распространения. Затем следует многоуровневая аннотация: тип манипуляции (аудио, визуал, текст), жанр и цель сообщения, стратегия убеждения или пропаганды, маркеры синтетичности. Методически целесообразно применять полуавтоматические разметочные среды с участием экспертов, привлекать LLM как ассистентов-аннотаторов с обязательной валидацией экспертами и разрабатывать метрики «степени синтетичности». Исследовательскими вопросами, которые неизбежно возникают при построении подобных корпусов, являются репрезентативность и сбалансированность корпуса по жанрам, языкам и каналам распространения, определение и классификация признаков пропагандистского контента и дипфейков, а также способы интеграции корпусов в сетевой анализ для купирования распространения манипулятивного контента.

Кейсы, подобные парламентским выборам в Индии, характеризующиеся массовым использованием ИИ, будут накапливаться и создавать «естественный эксперимент» для прикладной лингвистики. Комплексная разработка обозначенных направлений междисциплинарными коллективами – от корпусных и мультимодальных подходов до каузального моделирования и этического аудита алгоритмов, на наш взгляд, позволит не только описать новые формы политической речи, но и выработать устойчивые критерии ее анализа в эпоху генеративных технологий.

ЛИТЕРАТУРА

1. Dhanuraj D., Harilal S., Solomon N. Generative AI and its influence on India's 2024 elections: prospects and challenges in the democratic process : policy paper [Electronic resource]. URL: https://www.freiheit.org/sites/default/files/2025-01/a4_policy-paper_ai-on-indias-2024-electons_en-4.pdf (date of access: 27.07.2025).
2. Mitchell M., Wu S., Zaldivar A., Barnes P., Vasserman L., Hutchinson B., Spitzer E., Raji I.D., Gebru T. Model Cards for Model Reporting // FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency, January 29–31, 2019, Atlanta, GA, USA. 2019. P. 220–229. DOI: <https://doi.org/10.1145/3287560.3287596>.
3. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Electronic resource]. 2019. URL: <https://arxiv.org/abs/1810.04805> (date of access: 27.07.2025).

4. Zhou Z., Elejalde E. Unveiling the silent majority: stance detection and characterization of passive users on social media using collaborative filtering and graph convolutional networks // EPJ Data Science. 2024. Vol. 13, № 28. DOI: <https://doi.org/10.1140/epjds/s13688-024-00469-y>.
5. Ortiz-Garces I., Govea J., Andrade R.O., Villegas-Ch W. Optimizing Chatbot Effectiveness through Advanced Syntactic Analysis: A Comprehensive Study in Natural Language Processing // Applied Sciences. 2024. Vol. 14, № 5. Article № 1737. [Electronic resource]. DOI: <https://doi.org/10.3390/app14051737> (date of access: 27.07.2025).
6. Xu Y., Hu L., Zhao J. et al. A survey on multilingual large language models: corpora, alignment, and bias // Frontiers of Computer Science. 2025. Vol. 19, № 1911362. DOI: <https://doi.org/10.1007/s11704-024-40579-4>.
7. Arnold C., Küpfer A. Alignment Helps Make the Most of Multimodal Data [Electronic resource] // arXiv. 2024. URL: <https://arxiv.org/abs/2405.08454> (date of access: 27.07.2025).
8. Argyle L.P., Busby E.C., Gubler J.R., Lyman A., Olcott J., Pond J., Wingate D. Testing theories of political persuasion using AI // Proceedings of the National Academy of Sciences of the United States of America. 2025. Vol. 122, № 18. Article e2412815122. [Electronic resource]. DOI: <https://doi.org/10.1073/pnas.2412815122> (date of access: 27.07.2025).
9. Rosenberg S. Labeling won't solve AI's problems // Axios [Electronic resource]. 02.11.2023. URL: <https://www.axios.com/2023/11/02/ai-chatgpt-labeling-watermarking-human-machine> (date of access: 27.07.2025).
10. Waldorp L. J., Schutzeichel F., de Jong P. Causal inference: Principles unifying experimental and observational accounts // Theory & Psychology. 2025. [Electronic resource]. DOI: <https://doi.org/10.1177/09593543251325506> (date of access: 27.07.2025).
11. Bhashini – National Language Translation Mission [Electronic resource]. URL: <https://bhashini.gov.in> (date of access: 27.07.2025).
12. Li J., Galley M., Brockett C., Gao J., Dolan B. A Persona-based Neural Conversation Model // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Vol. 1: Long Papers. 2016. P. 994–1003. DOI: <https://doi.org/10.18653/v1/P16-1094>.

УДК 811.581,37+811.581,35

Сяо И

преподаватель
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Xiao Yi

Lecturer
Belarusian State University
of Foreign Languages
Minsk, Belarus
xiao.eric@yandex.by

СТРУКТУРНО-СЕМАНТИЧЕСКИЕ МОДЕЛИ ОБРАЗОВАНИЯ КИТАЙСКИХ ИДЕОГРАММ, ИМЕНУЮЩИХ АБСТРАКТНЫЕ ПОНЯТИЯ: К РАЗРАБОТКЕ БАЗЫ ЗНАНИЙ ДЛЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В данной статье с целью последующего создания базы знаний ИИ анализируются структурно-семантические модели образования 82 идеограмм, обозначающих абстрактные понятия в китайской письменности. Выделены три основные семантические категории идеограмм – абстрактных понятий: 1) государственно-общественные абстракции (73,1 %); 2) религиозно-ритуальные абстракции (15,9 %); 3) эмоционально-интеллектуальные абстракции (11,0 %). Определено, что идеограммы – абстрактные понятия в китайской письменности образуются путем соединения семантических компонентов со следующими значениями: 1) фигура человека и части его тела (人 ‘человек’, 心 ‘сердце’); 2) инструменты власти (矢 ‘стрела’, 戈 ‘алебарда’); 3) ритуальные объекты (祭 ‘алтарь’, 鼎 ‘треножник’). Работа вносит вклад в понимание механизмов знакообразования в иероглифике, подчеркивая роль культурно-исторического контекста в формировании идеограмм-абстрактных понятий. Она является основой для создания баз знаний для искусственного интеллекта.

Ключевые слова: идеограмма; субъект; иероглиф; смысловой компонент; структурно-семантическая модель; китайский язык.

STRUCTURAL-SEMANTIC MODELS OF CHINESE CHARACTERS DENOTING ABSTRACT CONCEPTS: TOWARD THE DEVELOPMENT OF KNOWLEDGE BASES FOR ARTIFICIAL INTELLIGENCE

This article analyzes structural-semantic models of 82 Chinese characters representing abstract concepts to inform the creation of AI knowledge bases. Three primary semantic categories of abstract-character characters are identified: 1) Socio-political abstractions (73,1 %); 2) Religious-ritual abstractions (15,9 %); 3) Emotional-intellectual abstractions (11,0 %). The study reveals that Chinese abstract characters are formed by combining semantic components with the following meanings: 1) Human figures and body parts (e.g., 人 ‘person’, 心 ‘heart’); 2) Instruments of authority (e.g., 矢 ‘arrow’, 戈 ‘halberd’); 3) Ritual objects (e.g., 祭 ‘altar’, 鼎 ‘tripod’). This work contributes to understanding hieroglyphic sign-formation mechanisms, emphasizing the role of cultural-historical context in shaping abstract-character semantics. It provides a foundation for developing AI knowledge bases.

Key words: ideogram; subject; logograph; semantic component; structural-semantic model; Chinese language.

Китайская иероглифика, будучи графической письменной системой с уникальной структурной сложностью, занимает особое место в мировой лингвистике как исследовательский объект. Ее ценность обусловлена многомерной репрезентацией не только языковых явлений, но и философских концептов, социально-исторических трансформаций и этнокультурных параметров эволюции древневосточной цивилизационной модели. Идеограмма – это иероглифический знак, который образуется через соединение только смысловых компонентов без включения какой-либо фонетической части. Идеограммы китайской иероглифической письменности формируются тысячелетиями в виде знаковой системы с внутренними особыми принципами организации. Они продолжают привлекать внимание не только лингвистов, но и ученых, работающих в области искусственного интеллекта. Понимание механизмов порождения китайских идеограмм, особенности такой сложной семантической области, как абстрактные понятия, позволит научить не только изучающих китайский язык, но и программное обеспечение правильно интерпретировать и, при необходимости, породить китайские знаки. Целью данной статьи, соответственно, является разработка структурно-семантических моделей соединения китайских иероглифов-идеограмм, которые обозначают абстрактные понятия различного типа.

Материалом исследования послужили 82 идеограммы, именующие абстрактные понятия, отобранные путем сплошной выборки из наиболее объемных и авторитетных словарей китайского языка: «汉语大字典» (Большой словарь китайских иероглифов) и «说文解字» (Объяснение простых и толкование сложных знаков) [1; 2]. В результате анализа отобранных идеограмм нами выделены три семантические группы знаков, которые создаются в китайском языке по типологии идеограмм. В каждой семантической группе идеограмм нами выявлены структурно-семантические модели их образования, т. е. те смысловые компоненты, которые входят в их состав, и как они комбинируются для создания того или иного знака.

Китайские идеограммы, обозначающие различные абстрактные понятия, представлены в виде трех семантических категорий знаков: 1) идеограммы, связанные с государством и обществом, составившие 73,1 % от общего числа отобранных единиц; 2) идеограммы, связанные с религией, составившие 15,9 %; 3) идеограммы, выражающие различные эмоции и мышление (11,0 %).

Наибольшую долю среди идеограмм, обозначающих абстрактные понятия, составили идеограммы, связанные с государством и обществом (73,1 %). Данные идеограммы создаются в китайской письменности по 13 структурно-семантическим моделям, количественная представленность которых в группе указана в таблице 1.

Т а б л и ц а 1

Количественная представленность структурно-семантических моделей
идеограмм, связанных с государством и обществом

Структурно-семантическая модель	Количество, %
«человек /часть тела» + «государство и общество»	20,0 %
«производства и бытия» + «производства и бытия»	20,0 %
«человек /часть тела» + «предмет природы»	11,7 %
«человек /часть тела» + «производства и бытия»	10,0 %
«человек /часть тела» + «Человек /часть тела»	8,3 %
«человек /часть тела» + «абстрактная сущность»	6,7 %
«государство и общество» + «предмет природы»	6,7 %
«производства и бытия» + «зоонимы»	5,0 %
«человек /часть тела» + «зоонимы»	3,3 %
«производства и бытия» + «зоонимы»	3,3 %
«производства и бытия» + «абстрактная сущность»	1,7 %
«государство и общество» + «государство и общество»	1,7 %
«абстрактная сущность» + «абстрактная сущность»	1,7 %

Наиболее распространенная структурно-семантическая модель позволяет создавать идеограммы-абстракции государства и общества путем объединения, с одной стороны, семантических компонентов, обозначающих человека или его часть, с другой стороны, семантических компонентов со значением понятий, связанных с государством и обществом. Например, идеограмма 侯 ‘Хоу (второй из пяти рангов шанской и чжоуской аристократии)’ складывается из смысловых компонентов 亻 ‘человек’, 夨 ‘человек на высоте’ и 矢 ‘стрела’, так как ранг Хоу – высокий уровень аристократии, аристократы этого ранга владели территорией вдоль границы, которую обязаны были оборонять [2; 3].

Высокочастотна структурно-семантическая модель, заключающаяся в редупликация смысловых компонентов, выражающих объекты для производства и бытия. Например, идеограмма 則 ‘закон, принцип’ имеет первоначальную форму написания 𠂔, включая такие семантические конститuenty, как 鼎 ‘емкость-треножник или четырехножник’ и 刀 ‘нож’, поскольку древние китайцы вырезали тексты законов для утверждения [3; 5].

Третья модель позволяет создавать идеограммы путем объединения, с одной стороны, семантических компонентов, обозначающих человека или часть его тела, с другой стороны, семантического компонента со значением природных предметов и явлений. Например, идеограмма 姓 ‘фамилия

(по материнской линии)’ создана из смысловых компонентов 女 ‘женщина’ и 生 ‘появление жизни’, так как первые китайские фамилии произошли от древнего матрилинейного общества.

Четвертая структурно-семантическая модель составляется из семантических компонентов, выражающих понятия человека или части его тела и семантического компонента, обозначающего предмет производства или бытия. Например, 守 ‘должность чиновника’ – идеограмма, в состав которой включаются смысловые компоненты 扌 ‘точка биения пульса, цунь (мера длины)’ и 宀 ‘крыша здания’. В понимании древних китайцев компонент 扌 ‘точка биения пульса, цунь (мера длины)’ представляет собой репрезентант инструмента для выражения понятий, связанных с правилами или порядками.

Пятая структурно-семантическая модель, по которой создаются идеограммы, связанные с государством и обществом, это редупликация смысловых компонентов, выражающих человека и части его тела. Например, 位 ‘позиция во властях’ – идеограмма структуре которой содержатся следующие смысловые компоненты: 亻 ‘человек’, 立 ‘смотрящий человек, видеть’. Сочетание данных компонентов дало древним китайцам понятия «позиция во властях», так как письменный знак 亻 ‘человек’ первоначально мог выражать социальное понятие «гражданин», т. е. человек с политическими правами [2; 4].

Шестая выделенная нами структурно-семантическая модель построения идеограмм представляет собой объединение семантического компонента со значением человека и части его тела и семантического компонента, выражающего количество. Например, идеограмма 丈 ‘чжан (китайская сажень, равна 3,33 метра)’ создана из смысловых компонентов 又 ‘правая рука’ и 十 ‘десять’, так как по древним правилам Китая единица «чжан» – это длина 10 рук [4, с. 165].

Седьмая разработанная нами структурно-семантическая модель позволяет образовать идеограмму с помощью таких смысловых компонентов, как «обозначение государства и общества» и «обозначение природного предмета». Например, 域 ‘государственная территория, пределы страны’ составлена из смысловых компонентов 或 ‘страна, государство’ и 土 ‘земля’.

Наименьшее распространение имеют следующие модели идеограмм, обозначающих абстрактные понятия, связанные с государственными делами и общественно-социальной жизнью: 8) «производство и бытие» + «зооним», например, 寶 ‘драгоценность’ = 宀 ‘крыша здания’ + 王(玉) ‘нефрит’ + 缶 ‘амфора’ + 貝 ‘ракушка’; 9) «человек/часть тела» + «зооним», например, 隻 ‘правила, совещание’ = 又 ‘правая рука’ + 萑 ‘сова’; 10) «производство и бытие» + «предмет природы», например, 災 ‘катастрофа’ = 火 ‘огонь’ + 巛 ‘река’; 11) «производство и бытие» + «абстрактная сущность» например, 罪 ‘преступление, нарушение закона’ = 罒 ‘сетка’ + 非 ‘вопреки, ошибка’;

12) «государство и общество» + «государство и общество» например, 族 ‘род, клан’ = 旂 ‘флаг’ + 矢 ‘стрела’; «абстрактная сущность» + «абстрактная сущность» например, 章 ‘часть музыкального произведения’ = 音 ‘музыка’ + 十 ‘десять’.

Не менее широкую по объему группу составили идеограммы, которые обозначают абстрактные понятия, связанные с религией (15,9 % от общего числа идеограмм-абстрактных понятий). Идеограммы этой семантической категории знаков созданы по 7 семантическим моделям, количественная представленность которых указана в таблице 2.

Т а б л и ц а 2

Количественная представленность структурно-семантических моделей идеограмм со значением абстрактных понятий, связанных с религией

Структурно-семантическая модель	Количество, %
«государство и общество» + «государство и общество»	23,1 %
«государство и обществом + «предмет природы»	23,1 %
«человек / часть тела» + «Человек / часть тела»	15,4 %
«человек / часть тела» + «предмет природы»	15,4 %
«человек / часть тела» + «производства и бытия»	7,7 %
«государство и общество» + «производства и бытия»	7,7 %
«предмет природы» + «предмет природы»	7,7 %

Наиболее распространенная модель позволяет образовать идеограмму, связанную с религией, путем редупликации семантических компонентов, выражающих понятия, связанные с государством и обществом. Например, 禮 ‘ритуал, этикет для уважения божеств и предков’ представляет собой идеограмму, куда включаются смысловые компоненты 礻 ‘алтарь’ и 豐 ‘посуда для жертвоприношения’. Письменный знак 礻 ‘алтарь’ в понимании древних китайцев может представлять собой инструмент для обозначения понятий, связанных с верованием и ритуалом [3; 4].

Высокочастотны также объединения, с одной стороны, семантических компонентов, обозначающих государство и общество, с другой стороны, семантического компонента со значением природных предметов и появлений. Например, идеограмма 社 ‘божество земли’ включает такие семантические конститuenty, как 礻 ‘алтарь’ и 土 ‘земля’.

Третья модель позволяет создавать идеограммы путем редупликации семантических компонентов, выражающих понятия, связанные с человеком и

частью его тела. Например, 𡗗 ‘молитвенная церемония с участием жреца или представителя божества’ – идеограмма, в составе которой содержатся смысловые компоненты 巫 ‘жрец’, 𠂇 ‘два рта, гомон’, и 𠂇 ‘пара рук’. Компонент 巫 ‘жрец’ представляет собой репрезентант субъекта производимого действия, а компоненты 𠂇 ‘два рта, гомон’ и 𠂇 ‘пара рук’ – инструменты.

Четвертая структурно-семантическая модель составляется из семантических компонентов, выражающих понятие человека или части его тела, и семантического компонента предмета природы. Например, идеограмма 仙 ‘небожитель, бессмертный’ создана из смысловых компонентов 亻 ‘человек’ и 山 ‘гора’, так как по китайской древней мифологии небожитель и бессмертный человек живут на высокой горе.

Наименьшее распространение имеют следующие модели идеограмм, обозначающих религиозные понятия: 5) «человек / часть тела» + «производство и бытие», например, 𡗗 ‘оборотень’ = 鬼 ‘дух’ + 𠂇 ‘рисунок’; 6) «государство и общество» + «производство и бытие», например, 福 ‘счастье, везение’ = 礻 ‘алтарь’ + 畐 ‘наполненная емкость для урожая или вина’; 7) «предмет природы» + «предмет природы», например, 坤 ‘энергия земли’ = 土 ‘земля’ + 申 ‘гром и молния’.

Наименьшую долю составили идеограммы, обозначающие различные эмоции и мышление человека (11,0 % от общего числа идеограмм, имеющих абстрактные понятия). Идеограммы этого семантического типа создаются в китайской письменности согласно 6 структурно-семантическим моделям. Количественное распределение единиц, которые могут быть отнесены к той или иной модели, представлено в таблице 3.

Т а б л и ц а 3

Количественная представленность структурно-семантических моделей, по которым образуются идеограммы, обозначающие различные эмоции и мышление человека в китайском языке

Структурно-семантическая модель	Количество, %
«Человек / часть тела» + «человек / часть тела»	22,2 %
«Человек / часть тела» + «абстрактная сущность»	22,2 %
«Человек / часть тела» + «предмет природы»	22,2 %
«Производства и бытия» + «предмет природы»	11,1 %
«Предмет природы» + «предмет природы»	11,1 %
«Человек / часть тела» + «государство и общество»	11,1 %

Структурно-семантическая модель, по которой чаще образуются знаки подобного типа, представляет собой редупликацию семантических компонентов, выражающих значение человека или его частей. Например, идеограмма 色 ‘настроение; выражение лица’ образована в результате соединения смысловых компонентов 亻 ‘человек’ и 邑 ‘человек на коленях’.

Высокочастнота структурно-семантическая модель, представляющая собой объединение, с одной стороны, семантических компонентов, обозначающих человека или часть его тела, с другой стороны, семантического компонента со значением абстрактных понятий. Например, идеограмма 意 ‘мысль, воля’ включает такие семантические конститuenty, как 心 ‘сердце’ и 音 ‘голос, музыка’, т. к. в понимании древних китайцев сердце — это место, где душа находится, является главным органом ментальной деятельности [2; 4].

Третья структурно-семантическая модель составляется из семантических компонентов, выражающих человека или части его тела и семантического компонента — предмета природы. Например, 厄 ‘переживание, мучение’ — идеограмма, в чей состав включаются смысловые компоненты 卩 ‘сидящий человек’ и 厂 ‘пещера на обрыве’.

Наименьшее распространение имеют следующие модели идеограмм, обозначающих абстрактные понятия, связанные с эмоциями и мышлением: 4) «производство и бытие» + «предмет природы», например, 夢 ‘сон’ = 苜 ‘подножный корм’ + 冃 ‘покрытие’ + 夕 ‘луна вечером’; 5) «предмет природы» + «предмет природы», например, 昊 ‘великое небо, небосвод’ = 日 ‘солнце’ + 天 ‘небо’; 6) «человек / часть тела» + «государство и общество», например, 詞 ‘слово, речь’ = 言 ‘рот говорящий’ + 司 ‘управление’.

Таким образом, анализ структурно-семантических моделей образования идеограмм, обозначающих абстрактные понятия различного типа в китайской письменности, показал, что графическое составление данных знаков обусловлено, прежде всего, включением в состав идеограмм семантических компонентов, которые выражают конкретные сущности, к которым логически могут быть отнесены описываемые идеограммами абстрактные понятия. Данными категориями предстают обозначения лиц или организма человека (亻 ‘человек’, 心 ‘сердце’, 女 ‘женщина’, 邑 ‘человек на коленях’, 口 ‘рот’ и др.) и объектов государственных и общественных дел (礻 (示) ‘алтарь’, 矢 ‘стрела’, 豐 ‘посуда для жертвоприношения’, 旗 ‘флаг’, 戈 ‘оружие’ и др.). Графемы, репрезентирующие эти категории модели мира, дополняются смысловыми компонентами, обозначающими важные для носителей китайского языка признаки, исторически связанные с целым рядом национально-культурных особенностей, широкая вариативность которых приводит к сосуществованию большого разнообразия знакообразовательных

моделей конструирования китайских идеограмм, описанных в данной работе. Среди дополнительных признаков наиболее часто используются смысловые компоненты, обозначающие объекты производства и бытия, объекты и явления живой природы.

ЛИТЕРАТУРА

1. 许慎. 说文解字 / 许慎, 张三夕, 刘果。 – 岳麓 : 岳麓书社, 2006. – 540页。 = Сюй, Шэнь. Объяснение простых и толкование сложных знаков (Шовэньцзецзы) / Сюй Шэнь, ЧжанСаньси, ЛюГо. Юэлу : Изд-во в Юэлу, 2006. 537 с.
2. 汉语大字典 / 汉语大字典编委会: 徐中舒、李格非、赵振铎等, 学术顾问组: 于省吾、王力等。 – 武汉 : 崇文书局, 2010. – 5727 页。 = Большой словарь китайских иероглифов / редкол. : СюйЧжоншу, Ли Гэфэй, ЧжаоЧжэньдо [и др.]. Ухань : Чунвэнь, 2010. 5727 с.
3. 黄天树. 说文解字通论。 – 北京: 北京大学出版社, 2014. – 361页。 = Хуан, Тяньшу. Общее суждение «Шовэньцзецзы» / Хуан Тяньшу. Пекин : Изд-во Пекинского ун-та, 2014. 361 с.
4. 李学勤. 字源。 – 天津: 天津古籍出版社, 2013. – 1420页。 = Ли, Сюэцин. Этимология иероглифов / Ли Сюэцин. Тяньцзинь : Старинные книги Тяньцзиня, 2013. 1420 с.
5. 谷衍奎. 汉字源流字典。 – 北京 : 语文出版社, 2008. – 2155页。 = Гу, Янькуй. Словарь этимологии китайских иероглифов / ГуЯнькуй. Пекин : Изд-во китайского языка и литературы, 2008. 2155 с.

УДК 81'42

Тулцева Анастасия Эдуардовна
старший преподаватель
Московский городской
педагогический университет
г. Москва, Россия

Anastasia Tultseva
senior lecturer
Moscow City University
Moscow, Russia
tulcevae@mgpu.ru

DEEPSEEK В НЕМЕЦКОМ МЕДИАПРОСТРАНСТВЕ: ОБРАЗЫ, НАРРАТИВЫ И ВЫЗОВЫ ВНЕДРЕНИЯ

Статья исследует лингвистическое конструирование образа китайского ИИ-ассистента DeepSeek в ведущих немецких медиа. На основе дискурс-анализа публикаций в немецких медиаисточниках (ARD, ZDF, Handelsblatt, RBB) выявляются доминирующие нарративы, языковые стратегии фрейминга и ключевые проблемные поля, связанные с возможностью внедрения ресурса в немецкоязычной среде. Основное внимание уделяется роли лингвистических средств в актуализации проблем доверия (обусловленного происхождением), соответствия регуляторным нормам (GDPR, AI Act) и культурно-языковой адаптации. Результаты показывают, что медийный дискурс формирует преимущественно проблемно-ориентированный образ DeepSeek, где лингвистические маркеры риска и неопределенности преобладают над описанием функционального потенциала, создавая существенные коммуникативные барьеры для его восприятия и принятия. Целью немецкого медийного дискурса является в основном создание негативного образа данного ИИ-ассистента у немецкоязычного пользователя. Большинство ведущих немецких СМИ подвергают сомнению необходимости его использования на территории немецкоязычного пространства.

Ключевые слова: DeepSeek; ИИ-ассистент; немецкие медиа; медийный дискурс; проблема доверия; негативный образ.

DEEPSEEK IN THE GERMAN MEDIA SPACE: IMAGES, NARRATIVES AND CHALLENGES OF IMPLEMENTATION

The article examines the linguistic construction of the image of the Chinese AI assistant DeepSeek in leading German media. Based on the discourse analysis of publications in German media sources (ARD, ZDF, Handelsblatt, RBB), dominant narratives, linguistic framing strategies and key problem areas related to the possibility of introducing the resource in the German-speaking environment are identified. The main attention is paid to the role of linguistic means in actualizing the problems of trust (due to origin), compliance with regulatory standards (GDPR, AI Act) and cultural and linguistic adaptation. The results show that the media discourse forms a predominantly problem-oriented image of DeepSeek, where linguistic markers of risk and uncertainty prevail over the description of the functional potential, creating significant communicative barriers to its perception and acceptance. The aim of the German media discourse is mainly to create a negative image of this AI assistant among German-speaking users. Most leading German media question the need for its use in the German-speaking area.

Key words: DeepSeek; AI assistant; German media; media discourse; trust issue; negative image.

Выход на международные рынки инновационных технологий, особенно ИИ-ассистентов на основе больших языковых моделей, сопровождается

сложным процессом их медийного и общественного осмысления. Китайский проект DeepSeek, позиционирующийся как мощный и бесплатный конкурент ChatGPT, представляет особый интерес для лингвистического исследования в силу своей новизны, культурно-географического происхождения и амбиций глобального присутствия. Целью данной статьи является лингвистический анализ репрезентации DeepSeek в немецких медиа, направленный на выявление:

- доминирующих образов и нарративов, конструируемых языковыми средствами;
- лингвистических стратегий фрейминга проблем использования и внедрения;
- ключевых дискурсивных барьеров, связанных с доверием, регуляцией и культурно-языковой спецификой.

Актуальность исследования обусловлена, во-первых, ролью медиадискурса как первичного источника формирования представлений о новых технологиях для широкой аудитории и профессиональных сообществ, во-вторых, спецификой немецкого медиапространства с его высокими требованиями к конфиденциальности и критическим отношением к технологиям из неевропейских юрисдикций, а также недостатком лингвистически ориентированных исследований восприятия конкретных больших языковых моделей, особенно «незападного» происхождения, в национальных медиа.

Не вызывает сомнения, что современные ИИ-ассистенты стали в том числе неотъемлемой частью современного языкового образования [1, с. 110]. Сгенерированные ими тексты в последнее десятилетие находят свое отражение в разных сферах современного общества, а особенности создания и восприятия таких текстов представляет собой отдельный интерес для изучения [2]; [4]. Изучающие иностранный язык активно используют различные медиаресурсы, в том числе в виде новостных мобильных приложений [5, с. 141], при создании информационного ресурса которых в ряде стран может быть использован в том числе ИИ-ассистент DeepSeek.

При лингвистическом анализе медиарепрезентации DeepSeek в немецких СМИ следует упомянуть следующие ключевые концепции:

- концепция критического дискурс-анализа Н. Фэрклафа, согласно которой дискурс как социальная практика определяет медиатексты не просто как отражающие реальность, но активно конструирующие ее через язык, воспроизводя или оспаривая властные отношения и социальные нормы [11]. Медийный дискурс о DeepSeek формирует его социальную «реальность» для немецкой аудитории;

- теория фреймов Р. Энтмана, где отбор и акцентирование определенных аспектов реальности осуществляется при игнорировании других [9]. Немецкие СМИ фреймируют DeepSeek через призму специфических проблемных областей (доверие, регуляция);

- конструирование образа технологии. Согласно З. Вайшенбергу, технологии не имеют имманентных значений; их смысл и ценность приписываются обществом через коммуникацию. Лингвистические средства – ключевой инструмент этого конструирования [14, S. 56]. Образ DeepSeek формируется в немецком медийном поле между его техническими характеристиками (*stark* ('мощный'), *kostenlos* ('бесплатный') и контекстуальными атрибутами (*chinesisch* ('китайский'), *rechtswidrig* ('антуправовой') [5];

- концепция Ж.-Ф. Лиотара [3] о кризисе метанарративов. В применении к медиа нарративы как устойчивые сюжетные схемы, организующие некий опыт и придающие смысл некоему новостному единству, восприятие новой технологии, в данном случае интернет-ассистента DeepSeek, часто укладывается в знакомые нарративы (*Konkurrent von ChatGPT* ('конкурент чата GPT') [16], *Verstoß gegen das Europäische Datenschutzgesetz* ('Нарушен против европейского закона о защите персональных данных') [12], *Wirtschaftliche Aufholjagd* ('Технологическая гонка') [8];

- регуляторный и культурный контекст Германии: нормы немецкого закона о защите персональных данных (*Datenschutzgrundverordnung*) [10], концепция *доверенного ИИ* – систем искусственного интеллекта, которые разработаны с учетом надежности, справедливости, прозрачности и ответственности – [10] и исторически обусловленная чувствительность к вопросам приватности и безопасности данных являются дискурсообразующими факторами, напрямую влияющими на языковые выборы.

Остановимся на лингвистическом конструировании образа и проблем китайского интернет-ассистента DeepSeek в немецких СМИ на основе выборки примеров из таких интернет-ресурсов, как ARD [6]; ZDF [13], [15]; Handelsblatt [7], [8] и RBB24 [12].

DeepSeek в немецких медиа-ресурсах преимущественно репрезентируется через следующие номинации и базовые атрибуты:

- происхождение и функция: *Der chinesische KI-Assistent DeepSeek, Chinas Antwort auf ChatGPT, Ein neuer Chatbot-Konkurrent aus China* [13]. Указание на Китай присутствует в подавляющем большинстве номинаций, становясь ключевым классификатором;

- ключевые атрибуты: *leistungsstark* ('производительный'), *kostenlos* ('бесплатный'), *mehrsprachig* ('многоязычный' – с частыми оговорками о качестве немецкого: *auch auf Deutsch, allerdings...*) [6], [7], [12];

- сравнительный фрейм используется почти исключительно в контексте сравнения с доминантами рынка: *eine kostenlose Alternative zu ChatGPT* ('бесплатная альтернатива чату GPT'), *kann mit Gemini mithalten* ('может соперничать с Gemini'), *noch nicht so ausgereift wie etablierte Dienste* ('еще не так зрел, как занимающие постоянное место на рынке сервисы'). Используются глаголы сравнения (*konkurrieren mit, mithalten mit, herausfordern*) и наречия степеней сравнения (*fast so gut wie, noch nicht so... wie*) [6], [8], [15].

При анализе упомянутых немецких медиа-ресурсов в контексте аналитических статей о регулировании ИИ, корпоративном использовании ИИ, цифровой безопасности выявлены также трех ключевых нарратива:

1. Китайский вызов в ИИ-гонке с использованием лексики: *Aufholjagd* ('гонка за лидером'), *Technologieführerschaft* ('технологическое лидерство'), *Ambitionen* ('амбиции'), *globale Präsenz* ('глобальное присутствие'), *neue Machtverhältnisse* ('новый баланс сил') [8]. Используются также метафоры гонки (*Wettlauf*, *Rennen*), битвы/соревнования (*Konkurrenzkampf*, *Herausforderung für den Westen*) [6].

2. Бесплатная альтернатива (чату GPT) с вопросом доверия (*Kostenlose Alternative mit Vertrauensfrage*) с использованием лексики *kostenlose Option* ('бесплатная опция'), *potenzielle Alternative* ('потенциальная альтернатива'), *aber...* ('но'), *großes Fragezeichen* ('большой вопросительный знак'), *Zweifel* ('сомнения'), *Vorbehalte* ('оговорки, предубеждения'), *Sicherheitsbedenken* ('опасения безопасности'), *Datenrisiko* ('риск защиты персональных данных'). При этом используются также следующие синтаксические контрастные конструкции: (*zwar leistungsstark, aber...* ('хотя высокопроизводительный, но'), *kostenlos, jedoch...* ('бесплатный, однако'), риторических вопросов *Kann man einem KI-Assistenten aus China vertrauen?* ('Можно ли доверять ИИ-ассистенту из Китая?'), модальных глаголов со значением возможности/необходимости/сомнения (*könnte, dürfte, müsste, sollte* + *Negation/вопрос*) [7], [8], [12].

3. Проблема защиты персональных данных с использованием лексики: *Daten-Souveränität* ('суверенитет данных'), *graue Zone* ('серая зона'), *rechtliche Risiken* ('юридические риски'), *Datenabfluss* ('утечка данных'), *chinesische Datenschutzgesetze* ('китайские законы о защите персональных данных'), *Zugriff durch Behörden* ('доступ властей') [6], [7], [12]. Нормы европейского закона о защите персональных данных [10] воспринимаются как некий абсолют, относительно DeepSeek применяются выражения неопределенности/сомнения с использованием наречий (*unklar* ('неясный'), *fraglich* ('спорный'), *zweifelhaft* ('сомнительный'), конструкции *Risiko besteht* ('возникает риск') [там же].

В проанализированных медиаресурсах с лингвистической точки зрения конкретизируется также актуализация проблемных полей на примере вызовов внедрения, где ключевым полем является доверие и безопасность, в котором используется, например, следующая лексика: *chinesische Herkunft* ('происхождение' – ключевой маркер), *staatliche Kontrolle* ('государственный контроль'), *Zugriffsmöglichkeiten Pekings* ('возможности доступа Пекина'), *Überwachungsgesetze* ('законы о слежке'), *Sicherheitslücke* ('брешь безопасности'), *'Trojanisches Pferd* ('Троянский конь' – метафора), *Risiko für Geschäftsgeheimnisse* ('риск для коммерческой тайны'), *Undurchsichtigkeit* ('непрозрачность'). Аргументация статей строится на отсылках к китайскому законодательству и принципиально иной модели регулирования

данных. Используются гипотетические сценарии (*könnte Daten weitergeben, dürfte verpflichtet sein* – ‘мог бы передавать данные’, ‘был бы обязан’) [7], [8], [12], [13].

Обязательно подчеркивается также соответствие немецким правовым нормам с помощью лексики: *DSGVO-Hürde* (‘препятствие немецкому закону о защите персональных данных’), *Rechtsunsicherheit* (‘правовая неопределенность’), *Anforderungen an Transparenz* (‘требования к прозрачности’), *Datenverarbeitung außerhalb der EU* (‘обработка данных вне ЕС’), *Recht auf Auskunft* (право на информацию). Частые вопросы о соответствии, использовании данного интернет-ассистента выражаются с помощью использования отрицательных или условных конструкций *bietet keine ausreichende Transparenz* (‘нет достаточной прозрачности’), *wenn Daten in China verarbeitet werden, dann...* (‘если данные обрабатываются в Китае, то’) [6], [7], [8], [12], [13].

Для оценки интернет-ассистента DeepSeek активно используются средства оценки посредством использования сравнительных конструкций с негативным оттенком (*schlechter als bei ChatGPT, noch nicht ausreichend für professionelle Anwendungen* – ‘хуже, чем у чата GPT’, ‘еще недостаточно для профессионального применения’) [12], причем делается акцент на недостатках немецкоязычной функциональности данного интернет-ассистента.

Проведенный анализ немецких интернет-медиаресурсов демонстрирует, что медийный дискурс о DeepSeek в Германии принципиально проблемно-центричен. Его лингвистическое конструирование характеризуется следующими факторами:

1. Доминированием атрибута происхождения: лексема ‘китайский’ выступает не просто дескриптором, но и ключевым дискурсивным маркером, автоматически активирующим фреймы риска, недоверия и регуляторных сложностей.

2. Преобладанием нарратива *доверия/безопасности*: языковые средства (вопросы, контрасты, гипотетические сценарии, негативно окрашенная лексика) систематически актуализируют сомнения в надежности и прозрачности платформы, связывая их с ее национальной принадлежностью. Нарратив *альтернативы* неизменно сопровождается оговорками.

3. Языковой легитимацией регуляторных барьеров: термины DSGVO [10] используются как нормативные ярлыки, а лингвистические конструкции (вопросы о соответствии, утверждения о неопределенности) изображают DeepSeek как объект, по умолчанию находящийся в конфликте с европейскими нормами, требуя от него доказательства обратного.

4. Языковым снижением функционального потенциала: описания технических возможностей (‘мощный’, ‘бесплатный’) часто принижаются немедленным указанием на языковые ограничения или погружаются в контекст проблем доверия/регулирования. Позитивные атрибуты не формируют самостоятельного сильного нарратива.

5. Сравнением как инструментом позиционирования риска: сравнение с ChatGPT служит не столько для оценки объективных характеристик, сколько для лингвистического подчеркивания уникального риска DeepSeek – его китайского происхождения и связанных с этим последствий. DeepSeek представляется в немецких СМИ как более рискованный аналог.

Таким образом, немецкие медиа, используя специфический набор лингвистических стратегий (фрейминг через происхождение, акцентирование гипотетических рисков, нормативное давление терминов DSGVO [10], контрастные конструкции, формируют дискурсивный барьер для DeepSeek. Этот барьер заключается не только в перечислении объективных проблем, но и в самом языке их описания, который создает атмосферу неопределенности, повышенного риска и необходимости особой осторожности, существенно затрудняя формирование позитивного образа и готовности к внедрению ресурса в профессиональной и бытовой среде Германии.

ЛИТЕРАТУРА

1. Гончарова В. А. Тенденции современного образования: от проблем к возможностям // Проблемы современного образования. 2021. № 5. С. 101–115. DOI 10.31862/2218-8711-2021-5-101-115. EDN ALNVOS.
2. Ионина А. А. Особенности создания и восприятия текста в Интернете // Вестник МГПУ. Серия: Филология. Теория языка. Языковое образование. 2010. № 2 (6). С. 33–39. EDN NBJTKH.
3. Лиотар Ж.-Ф. Состояние постмодерна. М.: Институт экспериментальной социологии, 1998. 159 с.
4. Петрова И. М., Иванова А. М., Никитина В. В. Современные цифровые технологии в лингвистических исследованиях: Учебное пособие для обучающихся по направлению «Лингвистика». М.: Общество с ограниченной ответственностью "Языки Народов Мира", 2022. 259 с. ISBN 978-5-6048046-8-1. EDNLDOROG.
5. Тульцева А. Э. Использование мобильных новостных приложений на занятиях по практике устной и письменной коммуникации на немецком языке в современном вузе (продвинутый этап обучения) // Германистика: филология и лингводидактика: СБОРНИК НАУЧНЫХ СТАТЕЙ. М.: Общество с ограниченной ответственностью «Языки Народов Мира», 2025. С. 140–148. EDN QIGEE0.
6. ARD: официальный сайт. DeepSeek droht Rauswurf aus deutschen App-Stores [Electronic resource]. URL: <https://www.tagesschau.de/inland/deepseek-datenschutz-102.html> (date of access: 14.07.2025)
7. Bomke L. Deepseek fordert OpenAI mit neuer KI-Version erneut heraus Talente [Electronic resource]. URL: <https://www.handelsblatt.com/technik/ki/kuenstliche-intelligenz-deepseek-fordert-openai-mit-neuer-ki-version-erneut-heraus/100116277.html> (date of access: 14.07.2025).

8. Benninghof M. Chinas harter Kampf gegen die USA um die besten KI-Talente [Electronic resource]. URL: <https://www.handelsblatt.com/technik/ki/kuenstliche-intelligenz-chinas-harter-kampf-gegen-die-usa-um-die-besten-ki-talente/100114623.html> (date of access: 14.07.2025)
9. Entman R. M. Framing: Toward clarification of a fractured paradigm // Journal of Communication, 1993. № 43(4). P. 51–58.
10. EU-Datenschutzverordnung [Electronic resource]. URL: <https://dsgvo-gesetz.de/> (date of access: 14.07.2025)
11. Fairclough N. Critical Discourse Analysis: The Critical Study of Language. London: Longman, 2010. 604 p.
12. RBB24 : официальный сайт. Berliner Datenschutzbeauftragte will Deepseek aus App-Stores verbannen [Electronic resource]. URL: <https://www.rbb24.de/politik/beitrag/2025/06/deepseek-appstore-datenschutzbeauftragte-berlin-meike-kamp-apple-google.html> (date of access: 14.07.2025).
13. Rieken S. DeepSeek wirbelt Börsen auf [Electronic resource]. URL: <https://www.zdfheute.de/video/heute-nachrichten/deepseek-ki-app-video-100.html> (date of access: 14.07.2025).
14. Weischenberg S. Nachrichtenjournalismus. Anleitungen und Qualitäts-Standarts für die Medienpraxis. Wiesbaden: Verlag für Sozialwissenschaften, 2013. 298 S.
15. ZDF: официальный сайт. Sicherheitsbedenken gegen DeepSeek [Electronic resource]. URL: <https://www.zdfheute.de/video/heute-nachrichten/videos/bsi-china-ki-deepseek-sicherheitsbedenken-video-100.html> (date of access: 14.07.2025).
16. ZDF: официальный сайт. Südkorea verbietet KI-App DeepSeek [Electronic resource]. URL: <https://www.zdfheute.de/politik/ausland/deepseek-verbot-sued-korea-ki-chatbot-100.html> (date of access: 14.07.2025).

УДК 004.8

Царева Виктория Олеговна

магистрант кафедры перевода
ФГБОУ ВО РГПУ им. А. И. Герцена
г. Санкт-Петербург, Россия

Viktoria Tsareva

Master's degree student
The Herzen State Pedagogical
University of Russia
Saint Petersburg, Russia
v2o2@yandex.ru

О ВОЗМОЖНОСТЯХ ГЕНЕРАЦИИ ЮМОРА СИСТЕМАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Генерация юмора с помощью искусственного интеллекта – это активно развивающаяся область на стыке лингвистики и машинного обучения. В данной статье представлен обзор современных методов создания юмористических текстов с использованием больших языковых моделей, таких как GPT. Рассматриваются ключевые преимущества ИИ, такие как скорость генерации текстов, грамматическая точность и способность воспроизводить стилистические особенности, а также основные ограничения: зависимость от качества обучающих данных, отсутствие контекстуального понимания и трудности передачи невербальных аспектов юмора. Особое внимание уделено этическим проблемам, включая предвзятость данных и риски генерации оскорбительного контента. Анализируются методы оценки юмора, такие как экспертные оценки и «коэффициент комического», а также перспективы дальнейшего развития технологий, включая мультимодальные решения. Статья подчеркивает потенциал ИИ как инструмента поддержки творчества, но отмечает, что создание по-настоящему оригинального юмора, учитывающего нюансы языкового, культурного, ситуативного и контекстуального порядка, остается сложной задачей.

Ключевые слова: искусственный интеллект; генерация юмора; большие языковые модели; обработка естественного языка; оценка юмора.

ON POSSIBILITIES OF HUMOUR GENERATION BY ARTIFICIAL INTELLIGENCE SYSTEMS

Artificial intelligence humour generation is an actively developing field at the intersection of linguistics and machine learning. This article presents an overview of current methods of humour text generation using large language models such as GPT. Key advantages of artificial intelligence, such as speed of text generation, grammatical accuracy and the ability to reproduce stylistic features, are discussed, as well as major limitations, such as dependence on the quality of training data, lack of contextual understanding and difficulties in conveying non-verbal aspects of humour. Particular attention is paid to ethical issues, including data bias and the risks of generating offensive content. Methods for assessing humour, such as expert judgement and the “humour quotient”, are analysed, as well as prospects for further technological developments, including multimodal solutions. The article highlights the potential of artificial intelligence as a tool to support creativity, but notes that creating truly original humour that is linguistically, culturally, situationally and contextually nuanced remains a challenge.

Key words: artificial intelligence; humour generation; large language models; natural language processing; humour evaluation.

Генерация юмора с помощью искусственного интеллекта (ИИ) – это активно развивающаяся область лингвистики и машинного обучения. Современные языковые модели, такие как GPT-3.5 и GPT-4, могут применяться для генерации шуток, языковой игры, стендап выступлений и других юмористических текстов. Однако способности ИИ остаются предметом дискуссий: насколько ИИ понимает юмор, может ли он создавать по-настоящему оригинальные и смешные тексты и каковы этические и технические ограничения таких систем?

Существует несколько основных исследовательских направлений в области цифровой лингвистики, изучающей юмор, – распознавание юмора, генерация юмора, а также различные модели оценки юмора.

Генерация юмора направлена на создание юмористических текстов – шуток, анекдотов, текстов стендап-выступлений, каламбуров. Современные методы генерации юмора можно разделить на три основные категории: шаблонные, нейросетевые и гибридные методы. Шаблонные методы основаны на заранее определённых структурах или шаблонах, которые заполняются словами или фразами, отобранными по определённым критериям. Методы, основанные на шаблонах, «используют лексическую замену, такую как использование синонимии, меронимии, гипонимии» [1, с. 17831], а также подбирают рифмующиеся слова [2, с. 6]. Заполнение шаблона позволяет создавать шутки, соответствующие определённому стилю или жанру, но ограниченность предопределёнными структурами снижает их оригинальность и не позволяет выйти за рамки определённого жанра [3, с. 4]. Нейросетевые методы используют современные языковые модели, такие как GPT, и глубокое обучение для генерации юмора без каких-либо структурных ограничений. Нейронные сети обучаются на больших объемах текстовых данных, содержащих юмористические тексты, и затем используют полученные знания для генерации новых шуток. Сгенерированные юмористические тексты оригинальнее шаблонных, однако некоторые исследователи отмечают, что процесс генерации юмора нейронными сетями непрозрачен для пользователей-людей, и пользователи «не могут понять, как модель придумала шутку, или дать ей указание генерировать шутки определённым образом, что может привести к созданию неуместного или оскорбительного юмора» [3, с. 1]. Некоторые исследователи создают собственные гибридные методы, объединяющие в себе преимущества шаблонных и нейронных подходов [1, с. 17826].

Большинство современных исследований сфокусировано на генерацию юмора с помощью больших языковых моделей (БЯМ). Генерация юмора искусственным интеллектом – одна из сложнейших задач в области обработки естественного языка. Сгенерированные юмористические тексты имеют ряд недостатков. Качество сгенерированного юмора напрямую зависит от качества данных, на которых обучается нейросеть. Нейросетям необходимы большие объёмы информации, которая в силу своего объёма может включать в себя неподходящие данные. Некачественно составленный корпус юмо-

ристических текстов, имеющий, например, повторяющиеся шуточки или неюмористические тексты, влияет на полученный результат [4, с. 5]. Также при создании корпуса для обучения исследователи сталкиваются с проблемой предвзятости данных, собираемых с онлайн-платформ, использующих, например, системы сбора материала на основе количества лайков как показателя успешности юмора. Алгоритмы продвижения контента создают ситуацию, когда определенные шутки становятся гораздо более заметными, чем другие, и собирают больше положительных оценок и комментариев. Не все онлайн-платформы обеспечивают свободный доступ к количеству просмотров какой-либо записи, из-за чего возникает трудность вычета процентного соотношения просмотров к лайкам или комментариям. Кроме того, на количество положительных оценок влияют факторы, не имеющие прямого отношения к качеству шутки, такие как популярность автора или использование определенных ключевых слов [5, с. 6]. Эти факторы приводят к тому, что корпус может оказаться нерепрезентативным и предвзятым. Кроме того, корпус данных для обучения может содержать оскорбительный юмор, который затем будет воспроизведен в сгенерированном тексте [4, с. 5]. Однако, системы модерации сгенерированных текстов могут значительно ограничивать тематику юмористических текстов. При генерации текстов стендап-выступлений, которые затрагивают табуированные в том или ином обществе темы и критику политических и социальных институтов, исследователи столкнулись с тем, что модели зачастую отказываются генерировать подобный контент, рассматривая запросы пользователей как «потенциально опасные» или «неинклюзивные». Недостаточно разнообразный материал для обучения нейросети также может привести к культурной и социальной предвзятости и воспроизведению стереотипов [7, с. 3]. Некоторые исследования также обнаружили, что модель GPT повторяет шуточки из обучающего корпуса, не внося никаких изменений [4, с. 5; 6, с. 325].

Другим недостатком сгенерированного юмора является отсутствие контекстуального понимания. Исследователи отмечают, что сгенерированный юмор «находится везде и нигде одновременно» [7, с. 10]. Юмор, предназначенный для узкой аудитории и основанный на личном опыте, «будет трудно создавать без источника информации о говорящих и их прошлом» [8, с. 16]. Успешный юмор также крайне зависим от культурного и географического контекста. ИИ, ограниченный только тем контекстом, который предоставлен в промпте, не способен полностью адаптировать материал под конкретную аудиторию или ситуацию, особенно при генерации стендап-выступлений. Использование различных методов промптинга и уточнение деталей не всегда улучшает полученные результаты [7, с. 7].

БЯМ создают письменные юмористические тексты, и одним из основных недостатков сгенерированного юмора является ограниченность в подаче. Шутки, представленные только в текстовом формате, часто оцениваются как менее смешные, чем те же шутки, но представленные в таких форматах, как

аудиовизуальные ролики, юмористические изображения или стендап-выступления [9, с. 10]. Подача (delivery) играет ключевую роль в создании комического эффекта. Подача включает в себя как вербальные, так и невербальные элементы – мимику, жесты и позы, интонацию, скорость и громкость речи, использование различных акцентов и изменение произношения, а также тайминг – координацию времени произнесения реплик, пауз и других элементов выступления, чтобы достичь оптимального комического эффекта [10, с. 379]. БЯМ могут сгенерировать качественные юмористические тексты, но комический эффект не будет таким же сильным. Исследователи D. Gorenz и N. Schwarz отмечают, что активное развитие технологий искусственного интеллекта может устранить этот недостаток, так как «уже разрабатываются и совершенствуются инструменты создания голоса, имитирующие человеческую речь, и инструменты генерации видео, что в будущем позволит интегрировать создание и подачу юмора в единый процесс» [9, с. 10].

Однако, современные исследования выделяют несколько преимуществ генерации юмора [7, с. 7; 9, с. 9]. В первую очередь, они отмечают скорость генерации текстов. P. Mirowskietal в своей работе наблюдает эффективность ИИ в качестве инструмента для быстрой генерации первоначального контента и структуры стендап-выступлений. Респонденты их исследования позитивно отзывались о скорости генерации и помощи в построении структуры, предоставляющий основу, которую затем можно самостоятельно доработать [7, с. 7]. Сгенерированные тексты грамматически верны и не содержат ошибок [6, с. 328]. Галлюцинации, то есть склонность больших языковых моделей выдавать ложную информацию за достоверную, не являются столь значительным недостатком при генерации юмористических текстов, поскольку, в отличие от журналистских текстов или научных статей, юмористические тексты не выполняют функции передачи фактической информации и используют многочисленные средства нарушения логики, поэтому они не обязаны соответствовать строгим стандартам точности [9, с. 1].

БЯМ также хорошо справляются с генерацией шаблонного юмора или юмористических текстов со строгой структурой [8, с. 15]. Исследование S. Kim и L. B. Chilton показало, что обученный ИИ успешно генерирует юмористические подписи к фотографиям, которые затем получили такую же высокую оценку, как и подписи, написанные человеком [8, с. 17]. Специально обученная языковая модель генерировала более качественные тексты, по сравнению с GPT. Модель успешно подбирала «аналогичные и соотносимые социальные ситуации», подстраивалась под аудиторию (использовала молодежный слэнг) и следовала структуре шуток [8, с. 15]. Работа Y. Chenetal также показывает эффективность специально обученных моделей, по сравнению с GTP-3 и GPT-3.5 [1, с. 17830]. Кроме того, ИИ способен успешно подражать стилю шуток, указанных в промпте. Работа D. Gorenz и N. Schwarz показала, что сатирические заголовки, созданные ИИ на основе корпуса заголовков сатирической газеты “The Onion”, были оценены как столь же

смешные, как и написанные профессиональными авторами. Это свидетельствует о способности ИИ анализировать и воспроизводить стилистические особенности, характерные для конкретного источника [9, с. 6]

Исследователи предлагают различные методы оценки качества юмора. В основном, исследователи смешивают сгенерированные и написанные человеком шутки, а затем предлагают оценить их, используя, например, шкалу Ликерта [5, с. 6]. В своей работе A. Mittalatal использует «коэффициент комического», который рассчитывается на основе интенсивности и продолжительности смеха, а также последующую оценку смеха человеком [11, с. 1]. В исследовании, проведенном N. Joshi, участники оценивали юмор, сгенерированный ИИ и созданный человеком. Результаты исследования показывают, что сгенерированные шутки получали одинаковую оценку с человеческими, а в случаях генерации игры слов шутки ИИ оценивались выше [12, с. 4]. Примечательно, что, когда создатель (ИИ или человек) шутки не был указан, оценки сгенерированного юмора значительно улучшались, что свидетельствует о наличии предвзятости, связанной с осведомленностью о происхождении текстов. Участники также с трудом отличали сгенерированные шутки от человеческих, несмотря на высокую уверенность в своих ответах, что подчеркивает прогресс ИИ в имитации человеческого юмора [12, с. 4]. Однако, исследования также показывают, что БЯМ генерируют однотипный юмор, который респонденты находят неоригинальным и несмешным [6, с. 332; 7, с. 11].

Генерация юмора с помощью искусственного интеллекта – это динамично развивающаяся область. Современные языковые модели способны создавать грамматически правильные и структурно сложные юмористические тексты, демонстрируя потенциал в качестве инструментов для творчества. Однако их способность генерировать оригинальный юмор, который бы полностью учитывал контекст, остается ограниченной. Ключевые проблемы включают зависимость от качества обучающих данных, отсутствие глубокого понимания культурных и социальных нюансов, а также трудности в передаче невербальных аспектов юмора. Несмотря на эти ограничения, ИИ может служить полезным инструментом для быстрой генерации идей и структурирования материала, особенно в шаблонных жанрах. Дальнейшее развитие технологий, включая улучшение моделей и интеграцию мультимодальных возможностей, может помочь преодолеть существующие барьеры. В то же время важно учитывать этические аспекты, такие как предотвращение предвзятости и контроль над генерацией потенциально оскорбительного контента, чтобы обеспечить ответственное использование этих технологий.

ЛИТЕРАТУРА

1. Chen Y. et al. Talk Funny! A Large-Scale Humor Response Dataset with Chain of-Humor Interpretation [Electronic resource] // Proceedings of the AAAI Conference on Artificial Intelligence. 2024. Vol. 38. № 16. P. 17826–17834. DOI: <https://doi.org/10.1609/aaai.v38i16.29736>.

2. Winters T. Computers Learning Humor Is No Joke. [Electronic resource] // Harvard Data Science Review. 2021. Vol. 3. № 2. P. 1–18. DOI: <https://doi.org/10.1162/99608f92.f13a2337>.
3. Chen Y., Shi B., Si M. Prompt to GPT-3: Step-by-Step Thinking Instructions for Humor Generation [Electronic resource] // arXiv preprint arXiv:2306.13195. 2023. P. 1–5. DOI: <https://doi.org/10.48550/arXiv.2306.13195>.
4. Faulkner R., Rex R., Ryan M. Knock Knock: Neural Joke Generation and Classification Final Report [Electronic resource] // Georgia Institute of Technology Technical Report. 2023. URL: <https://michryan.com/project/humor-bot/> (date of access: 18.06.2025).
5. Winters T., Nys V., De Schreye D. Automatic joke generation: Learning humor from examples [Electronic Resource] // Distributed, Ambient and Pervasive Interactions: Technologies and Contexts. DAPI 2018. Lecture Notes in Computer Science. 2018. Vol. 10922. P. 360–377. DOI: https://doi.org/10.1007/978-3-319-91131-1_28.
6. Jentzsch S., Kersting K. ChatGPT is fun, but it is not funny! Humor is still challenging Large Language Models [Electronic resource] // Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis. 2023. P. 325–340. DOI: <https://doi.org/10.48550/arXiv.2306.04563>.
7. Mirowski P. W., Love J., Mathewson K., Mohamed S. A Robot Walks Into a Bar: Can language models serve as creativity support tools for comedy? An evaluation of LLMs' humour alignment with comedians [Electronic resource] // Proceedings of the ACM Conference on Fairness, Accountability, and Transparency. 2024. P. 1622–1636. DOI: <https://doi.org/10.1145/3630106.365899>.
8. Kim S., Chilton L. B. AI Humor Generation: Cognitive, Social and Creative Skills for Effective Humor [Electronic resource] // arXiv preprint arXiv:2502.07981. 2025. P. 1–18. DOI: <https://doi.org/10.48550/arXiv.2502.07981>.
9. Gorenz D., Schwarz N. How funny is ChatGPT? A comparison of human- and A.I.-produced jokes [Electronic Resource] // PLoS ONE. 2024. Vol. 19. № 7. DOI: <https://doi.org/10.1371/journal.pone.0305364>.
10. Double O. Getting the Joke: The Inner Workings of Stand-up Comedy. London : Bloomsbury Publishing, 2014. 536 p.
11. Mittal A. et al. “So You Think You're Funny?": Rating the Humour Quotient in Standup Comedy [Electronic resource] // arXiv preprint arXiv:2110.12765. 2021. P. 1–7. DOI: <https://doi.org/10.48550/arXiv.2110.12765>.
12. Joshi N. N. Evaluating human perception and bias in AI-generated humor [Electronic resource] // Proceedings of the 1st Workshop on Computational Humor (CHum). 2025. P. 79–87. URL: <https://aclanthology.org/2025.chum-1.9/> (date of access: 18.06.2025).
13. Borji A. A categorical archive of chatgpt failures [Electronic resource] // arXiv preprint arXiv:2302.03494. 2023. P. 1–41. DOI: <https://doi.org/10.48550/arXiv.2302.03494>.

14. Avetisyan H., Safikhani P., Broneske D. Laughing Out Loud – Exploring AI-Generated and Human-Generated Humor [Electronic resource] // 4th International Conference on NLP & Artificial Intelligence Techniques (NLAI 2023). 2023. P. 59–76. DOI: <https://doi.org/10.5121/csit.2023.132406>.
15. Cowie R. Computational research and the case for taking humor seriously [Electronic resource] // HUMOR. 2023. Vol. 36. № 2. P. 207–223. DOI: <https://doi.org/10.1515/humor-2023-0021>.
16. Kalloniatis A., Adamidis P. Computational humor recognition: a systematic literature review [Electronic resource] // Artificial Intelligence Review. 2024. Vol. 58. P. 1–52. DOI: <https://doi.org/10.1007/s10462-024-11043-3>.
17. Barattieri di San Pietro C., Frau F., Mangiaterra V., Bambini V. The pragmatic profile of ChatGPT: Assessing the communicative skills of a conversational agent [Electronic resource] // SistemiIntelligenti. 2023. Vol. 35. № 2. P. 379–399. DOI: <https://doi.org/10.1422/108136>.

УДК 81-2:004

Шеховская Юлия Андреевна

кандидат филологических наук,
доцент кафедры иностранных
языков

Тюменское высшее военно-
инженерное командное училище
имени маршала инженерных войск
А. И. Прошлякова
г. Тюмень, Россия

Yulia Shekhovskaya

PhD in Philology,
Associate Professor, Department of
Foreign Languages
Tyumen Higher Military Engineering
Command School named after Marshal
of Engineering Troops A. I. Proshlyakov
Tyumen, Russia
y.a.shekhovskaya@gmail.com

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СОМАТИЧЕСКИЕ МЕТАФОРЫ В ПОЛИТИЧЕСКОМ ДИСКУРСЕ

В статье рассматриваются когнитивные механизмы использования соматических метафор и фразеологизмов в политическом дискурсе с точки зрения их манипулятивного потенциала и коммуникативной эффективности. Особое внимание уделяется современным достижениям в области искусственного интеллекта, включая технологии автоматического сбора, классификации и генерации идиоматических выражений. Анализируются примеры политической риторики, а также возможные сценарии применения нейросетей для оптимизации коммуникации и межкультурной адаптации.

Ключевые слова: соматическая метафора; политический дискурс; искусственный интеллект; когнитивная лингвистика; фразеологизм; NLP; генеративные модели.

ARTIFICIAL INTELLIGENCE AND SOMATIC METAPHORS IN POLITICAL DISCOURSE

The article explores the cognitive mechanisms behind the use of somatic metaphors and idioms in political discourse, emphasizing their manipulative potential and communicative effectiveness. Special attention is paid to current advancements in artificial intelligence, including technologies for automatic collection, classification, and generation of idiomatic expressions. The article provides examples from political rhetoric and outlines possible applications of neural networks for communication optimization and intercultural adaptation.

Key words: somatic metaphor; political discourse; artificial intelligence; cognitive linguistics; idiom; NLP; generative models.

Современная политическая коммуникация базируется не только на фактах, но и на образах, метафорах, культурных кодах. Среди наиболее мощных когнитивных инструментов воздействия – соматические фразеологизмы, то есть выражения, в которых используются наименования частей тела. Их образность и культурная обусловленность делают их особенно эффективными в процессе формирования общественного мнения.

С развитием технологий обработки естественного языка (NLP) и искусственного интеллекта (ИИ) открываются новые перспективы анализа, интерпретации и генерации подобных выражений, в том числе в рамках политических и медийных дискурсов.

Соматическая метафора отражает универсальные телесные и культурные переживания человека. Выражения типа «держат руку на пульсе», «иметь железные нервы», «выходить из себя» или «кивать на других» не только метафоричны, но и глубоко укоренены в менталитете, когнитивной структуре и коммуникативных привычках носителей языка.

Такие выражения обладают высокой семантической плотностью и эмоциональной насыщенностью, что делает их незаменимыми в публичной риторике. Политические тексты насыщены подобными конструкциями, способными воздействовать на восприятие аудитории на бессознательном уровне [1, с. 56].

Современные модели NLP позволяют осуществлять автоматический поиск соматических фразеологизмов в больших объемах текстов, таких как парламентские стенограммы, политические заявления, медиаархивы. Используются инструменты, построенные на архитектурах трансформеров (BERT, RoBERTa, GPT), способные учитывать контекст и семантические сдвиги.

На основе полученных данных формируются паттерны употребления, их частотность и тональность. Например, модель может выявить, что выражение «закрыть глаза» чаще используется в контексте оправдания бездействия, тогда как «сжать кулаки» – в контексте мобилизации.

ИИ позволяет строить когнитивные карты, в которых телесные метафоры связываются с политическими концептами: сила, контроль, страх, сопротивление. Такие карты помогают выявлять скрытые стратегии убеждения, оппозиции или подавления.

Нейросетевые генеративные модели, обученные на корпусах политических текстов, могут создавать сообщения с использованием соматических фразеологизмов. Возможно также их автоматическое преобразование при переводе – с учетом национальных метафорических норм. Так, англоязычное выражение “keep a stiffupperlip” может быть адаптировано как «держаться молодцом» в русском варианте [2, с. 6].

ИИ-системы уже используются для оценки эффективности идиом. Анализ комментариев в соцсетях, реакции на выступления позволяет оценить, какие выражения вызывают наибольший отклик. На основе этих данных выстраиваются рекомендации по их дальнейшему использованию.

Создаются обучающие модули для журналистов, политиков, PR-специалистов, в которых ИИ предлагает наиболее эффективные фразеологизмы в заданных условиях – например, «мягкие» выражения для дипломатии или «жесткие» для оппозиционных выступлений.

Метафорическая репрезентация тела может сильно варьироваться в зависимости от культуры. То, что воспринимается как выразительное и приемлемое в одном языке, может оказаться неуместным в другом. ИИ должен учитывать культурные особенности при генерации и переводе идиом.

Особую сложность представляет интерпретация амбивалентных или устаревших выражений, где телесные метафоры могут терять свою актуальность или понимание среди молодежной аудитории [4, с. 28].

Использование искусственного интеллекта в изучении соматических метафор представляет собой перспективное направление, сочетающее когнитивную лингвистику и вычислительные технологии. Внедрение таких моделей открывает новые горизонты для анализа политического дискурса, однако требует критического осмысления и систематического подхода.

Во-первых, при обучении нейросетевых моделей наблюдается зависимость от качества и объема исходных корпусов. Политические и медийные тексты, будучи источником идиом, часто подвержены жанровым ограничениям, стилистической однотипности и идеологическим рамкам. Это может привести к сужению спектра интерпретаций, снижению культурной вариативности и формированию предсказуемых, шаблонных метафорических конструкций.

Во-вторых, одной из ключевых задач остается проблема сохранения культурной специфики при генерации и переводе метафор. Например, выражение «иметь зуб на кого-либо» в русском языке обладает негативной коннотацией мести, в то время как дословный перевод на другие языки может быть непонятен или воспринят нейтрально. ИИ-модель, не обладающая знанием культурных коннотативных связей, рискует утратить семантическую глубину при адаптации подобных выражений.

В перспективе возможна разработка мультиагентных систем, способных к сопоставительному анализу языков и культур на основе параллельных корпусов, включающих семиотические и этнокультурные данные. Такие системы могут автоматически учитывать частотность, контекстуальную окраску и прагматическую нагрузку идиом, создавая более точные и уместные интерпретации [3, с. 41].

Технологическим прорывом может стать включение в анализ мультимодальных параметров: интонации, мимики, визуальных элементов, сопровождающих использование фразеологизмов. Это позволит создать когнитивные профили пользователей или ораторов, выявлять повторяющиеся паттерны риторического поведения, а также формировать персонализированные рекомендации по использованию метафор.

С точки зрения практической реализации ИИ в этой области возможно создание:

- обучающих платформ для специалистов в области политического PR, позволяющих встраивать соматические идиомы в выступления с учетом целевой аудитории и коммуникативной цели;
- онлайн-инструментов автоматического анализа и маркировки идиом в публичных речах с оценкой степени их воздействия;
- средств обратной связи, анализирующих реакции в реальном времени и предлагающих альтернативные фразеологические конструкции в зависимости от эмоционального фона [5, с. 36].

Следует отметить и потенциальные риски. Некорректное использование фразеологизмов, особенно в межкультурной среде, может привести к комму-

никативным сбоям, усилению стереотипов или даже формированию конфликтного дискурса. ИИ в таком случае должен включать механизмы этической фильтрации и контекстной переоценки.

Современные разработки в области искусственного интеллекта демонстрируют высокую степень интеграции с гуманитарными науками, в частности – с когнитивной лингвистикой, социальной психологией и культурологией. Использование ИИ в анализе соматических метафор способствует формированию новых научных подходов, позволяющих по-новому взглянуть на структуру коллективного сознания, механизмы массовой коммуникации и способы идеологического воздействия через язык.

Особое значение имеет возможность системного анализа фразеологических кластеров, включающих телесные образы. Такие кластеры не только отражают особенности мышления носителей языка, но и выступают индикаторами социокультурных трансформаций. Например, смещение акцента в использовании выражений от агрессивных телесных метафор («ударить по рукам», «жёсткая рука власти») к нейтральным или примирительным («подать руку», «держат баланс») может свидетельствовать о смене парадигмы общественного дискурса.

ИИ-технологии позволяют фиксировать подобные сдвиги в реальном времени, анализируя не только лексическую структуру, но и динамику развития языковых паттернов, их привязку к конкретным историческим или политическим событиям. Это открывает возможности для прогнозирования общественных настроений, изучения дискурсивных стратегий политических акторов и выявления латентных смыслов в публичной риторике.

Интеграция с социальными науками дает возможность построения комплексных моделей взаимодействия языка, мышления и поведения. Такие модели могут использоваться, например:

- при анализе электоральных предпочтений на основе речевых паттернов;
- в системах предиктивной аналитики для медиаплатформ и новостных агрегаторов;
- в мониторинге деструктивных речевых практик и экстремистской риторики с целью их нейтрализации.

Также возможно создание социолингвистических атласов на основе автоматического картирования частотности и прагматики употребления соматических фразеологизмов в различных регионах, социальных группах и профессиональных сообществах. Такие инструменты могут быть полезны при разработке образовательных и интеграционных программ, особенно в условиях мультикультурного общества.

Развитие цифровых гуманитарных наук (Digital Humanities) в данном контексте получает новое наполнение: язык как система фразеологических образов предстает не только объектом анализа, но и ключом к интерпретации

социокультурной реальности. Искусственный интеллект, взаимодействуя с методами лингвистического и культурного анализа, становится не просто инструментом, а полноценным участником исследовательского процесса.

Интеграция ИИ в анализ и производство соматических фразеологических единиц открывает новые горизонты для лингвистики, политической риторики и межкультурной коммуникации. Создание автоматизированных систем мониторинга, генерации и адаптации фразеологизмов делает политические послания более точными, выразительными и целенаправленными.

Использование когнитивного и культурного потенциала соматических метафор с помощью ИИ способствует формированию новой языковой экосистемы, в которой эмоциональность, образность и эффективность могут быть просчитаны, оптимизированы и адаптированы под конкретную аудиторию.

ЛИТЕРАТУРА

1. Тарасова М. В., Огородов М. К., Петрова М. Л. Виды диалогизма во французском газетном дискурсе // Филологические науки в МГИМО. 2019. № 4 (20). С. 56–63.
2. Семенов А. В., Семенова Т. Г. Реферирование и аннотирование на русском и китайском языках. М.: Шанс, 2019. 295 с.
3. Lakoff G., Johnson M. *Metaphors We Live By*. Chicago: University of Chicago Press, 2003. 256 p.
4. Shutova E. Design and Evaluation of Metaphor Processing Systems. *Computational Linguistics*. 2020. Vol. 46(2). P. 341–373.
5. Bender E., Koller A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *ACL*. 2020. P. 5185–5198.

УДК 81-11

Kimiaki Tokumaru

LLB

Oita, Japan

tokumaru@app.iij4u.or.jp

RESTORING SCIENTIFIC MEMORY: REGENERATIVE RECEPTION OF CONCEPTS VIA LARGE LANGUAGE MODELS AND NEURO-LOGICAL INTEGRATION

Scientific knowledge is not merely accumulated – it is biologically and logically regenerated. This paper proposes a model of *regenerative reception* for restoring fragmented scientific concepts, integrating three key components: (1) the biological substrate of conceptual memory, hypothesized to reside in cerebrospinal fluid (CSF)-resident B-cells; (2) the logical distinction between semantic and episodic memory in scientific versus everyday concept formation; and (3) the instrumental role of Large Language Models (LLMs) in detecting contradictions at the physical layer of language.

We argue that LLMs, while incapable of semantic integration, can reliably identify terminological and structural inconsistencies in scientific texts. These contradictions, once flagged, must be addressed through a human-led regenerative process grounded in primary sources and the logical architectures established by original discoverers. Interpretation is explicitly prohibited until full epistemic reconstruction is achieved. The result is not a static archive but a *regenerative ecology of scientific memory*, where meaning is not transmitted but reactivated within the biological consciousness of the learner.

Key words: regenerative reception; scientific memory; B-lymphocytes in CSF; contradiction detection; physical vs logical layer; large language models (LLMs), semantic reconstruction.

1. Introduction: Scientific Memory and Its Fragmentation

Modern science, despite its unparalleled achievements, suffers from a persistent epistemological problem: the fragmentation, distortion, and forgetting of scientific concepts across time and disciplines. Foundational ideas are often misunderstood, misrepresented, or reduced to simplified narratives that obscure the original logic and context of their discovery. This phenomenon, which we refer to as the erosion of “scientific memory”, poses a significant challenge to both knowledge continuity and interdisciplinary synthesis.

In this paper, we propose a regenerative framework for the reception of scientific concepts, supported by Large Language Models (LLMs) as tools for contradiction detection, and grounded in a biological and logical theory of meaning. Our method distinguishes two critical layers of conceptual existence:

The **physical layer**, where signals, symbols, and translations occur, yet no intrinsic meaning resides.

The logical layer, where meaning is integrated, verified, and regenerated through human reasoning grounded in biological and experiential foundations.

By leveraging the surface contradictions within the physical layer, LLMs can assist in identifying inconsistencies, overlaps, or gaps in the scientific record.

However, the actual reception and restoration of meaning must be performed by humans through a process we term regenerative reception—a systematic reconstruction grounded in original observations, primary records, and the bodily experience of discovery.

To this end, we argue for a new kind of science-assisted methodology, one that blends immuno-cognitive foundations, logical epistemology, and AI-powered contradiction analytics to restore the coherence and integrity of scientific knowledge.

2. The Biological Substrate of Conceptual Meaning

Meaning is not an abstract computation, nor merely a linguistic convention. It is a biologically sustained state, grounded in cellular processes that support memory, inference, and integration. In this context, we hypothesize that B lymphocytes residing in the cerebrospinal fluid (CSF) may play a pivotal role in the stabilization and differentiation of conceptual meaning.

Although B-cells are conventionally associated with immune defense, this perspective is historically conditioned by their discovery in peripheral blood. From an evolutionary standpoint, it is plausible that their original habitat was within the ventricular system of the brain, where they may have contributed primarily to cognitive and integrative functions long before being recruited into systemic immunity.

These B-cells may recognize, bind, and network symbolic or sensory patterns, forming a distributed system together with CSF-contacting neurons (CSF-cNs). This system may support the transformation of multimodal input into coherent semantic constructs. Whether B-cells directly process grammatical relations remains an open question; an alternative possibility is that CSF-cNs, through their antigenic and ciliary mechanisms, encode phonological vectors to which B-cells respond only at a semantic level.

The physiological features of B-cells – particularly their surface antibodies and the combinatorial diversity of complementarity-determining regions (CDRs) – provide a structural basis for scalable conceptual memory. Rather than imposing meaning from above, this system enables an emergent and embodied internal language network, capable of long-term semantic stability and integration across modalities.

Rather than emphasizing how B-cell-based systems might resist translation or reinterpretation, we note instead their inherently interdisciplinary and mobile nature. Situated within the CSF, free from rigid cortical localization, these cells can traverse cognitive boundaries much as scientific concepts ideally should. Indeed, the very notion of a concept that resists disciplinary translation invites skepticism: What architecture could ever enforce such isolation?

In sum, scientific concepts may not only reside in minds or texts, but in the immunological architecture that mediates signal-to-meaning transformation.

This distinction is further supported by how scientific and everyday concepts are biologically integrated. Everyday concepts tend to emerge inductively, synthesized from episodic memories through repeated lived experience. By

contrast, scientific concepts are deductively constructed through the systematic integration of semantic memory – encoded and managed by other B-cells – and must ideally remain independent from direct episodic anchoring. Their coherence depends on a logical, trans-episodic structure that permits abstraction, generalization, and critical reinterpretation. Thus, scientific concepts should not rely on individual episodes, but instead be stabilized through networks of meaning-bearing cells capable of supporting deductive consistency and conceptual autonomy. The restoration of scientific memory, therefore, requires us to examine not only historical records but also the cellular substrates through which meaning is biologically encoded and preserved.

3. The Logic of Concept Formation

The formation of concepts is not a neutral cognitive process; it is shaped by the epistemological posture of the thinker and the structural properties of the knowledge domain. In this chapter, we distinguish two contrasting modes of concept formation: one governed by inductive generalization from experience, and the other by deductive structuring of semantic relationships.

Everyday concepts, which dominate our intuitive, narrative, and social cognition, are typically formed through the aggregation of episodic memories. They are enriched by context, culturally variable, and prone to metaphor and emotional resonance. These concepts serve practical purposes but are inherently fluid and fragmentary.

In contrast, **scientific concepts** must transcend personal experience and cultural contingency. They require deductive rigor, internal consistency, and logical autonomy. They are not merely names attached to phenomena but are entities defined by their position within a formal network of relations. Crucially, scientific concepts are not invented afresh by each learner; they are re-constructed, using semantic memory structures stabilized through biological and social transmission. The internal language of scientific thought, as argued in the previous chapter, depends on meaning-bearing B-cell networks rather than episodic accumulation.

This difference gives rise to an essential asymmetry in their pedagogical treatment: while everyday concepts may be acquired through immersion and imitation, scientific concepts demand structured learning, formal definition, and abstraction from anecdotal grounding. Misunderstandings arise when scientific concepts are treated inductively – forced into the mold of lived familiarity rather than logical necessity.

Therefore, the process of regenerative reception must explicitly shift the cognitive mode of the learner from inductive to deductive, from contextually entangled to logically disciplined. This shift is not merely intellectual – it is neurobiological, requiring a reconfiguration of how meaning is stored, stabilized, and activated. It is here that the immune-based internal memory system offers a promising substrate for scientific reasoning, not only because of its biological persistence but also because of its capacity for logical closure.

4. Translation and AI: Bound to the Physical Layer

Scientific meaning is often lost in translation – not merely between languages, but across the boundary that separates symbolic representation from logical understanding. Translation, in its conventional form, operates entirely within the physical layer: converting one signal or symbol into another according to pre-defined mappings. It does not guarantee the preservation or restoration of meaning, especially when meaning resides in biologically encoded and logically structured networks.

Large Language Models (LLMs), despite their remarkable linguistic fluency, are similarly confined to this physical layer. They manipulate symbol sequences, infer patterns, and generate coherent text, but they lack access to the logical and biological processes through which meaning is authentically integrated. In essence, LLMs function as sophisticated translators – adept at replicating surface-level coherence without entering the domain of semantic truth.

This distinction becomes crucial when attempting to restore scientific memory. AI can assist in identifying contradictions, ambiguities, redundancies, and discontinuities within the symbolic record, but it cannot adjudicate meaning or recover the experiential and logical foundations upon which concepts were originally built. In fact, efforts to extract truth directly from LLMs without contextual guidance risk reinforcing surface-level biases and historically accumulated distortions.

To harness AI effectively, we must restrict its role to contradiction tagging within the physical layer. By identifying linguistic or structural mismatches – without interpreting them – LLMs can provide a useful map of problematic areas. These flagged regions can then be subjected to human-led regenerative reception, wherein the original conceptual architecture is reconstructed through reference to primary documents, recorded experiences, and the embodied reasoning of original discoverers.

Moreover, when records or memoirs are missing, such contradictions should be tagged accordingly – “insufficient primary source documentation” – to lower their epistemic priority and discourage speculative interpretation. This reinforces the principle that meaning cannot be inferred from signal alone; it must be traced back to its biological and historical genesis.

In summary, translation is signal-level work. Scientific concepts, by definition, do not exist in the vocabulary of any natural language; they are specialized terms created to escape the constraints of everyday speech. As such, they are fundamentally untranslatable in any traditional sense. The best practice, therefore, is either to leave them untranslated or to adopt faithful transliterations – just as “tsunami” has become a globally accepted term. LLMs can assist in this task not by interpreting, but by verifying whether translations are faithful and consistent.

Indeed, one major historical example of mistranslation is Chomsky’s rendering of Saussurean terms “langue” and “langage” both as “language” in English. This conflation arguably compromised an entire tradition of linguistic

distinction. Such mistranslations must be identified and flagged – ideally by LLMs operating strictly within the physical layer – so that future conceptual reconstructions are based on structurally intact terminology.

LLMs operate at this level, and their utility lies in the systematic exposure of contradiction and terminological distortion – not the resolution of conceptual meaning. It is the human learner – biologically equipped for semantic integration – who must undertake the work of restoration. LLMs operate at this level, and their utility lies in the systematic exposure of contradiction, not the resolution of conceptual meaning. It is the human learner – biologically equipped for semantic integration – who must undertake the work of restoration.

5. The Prohibition of Interpretation: Regenerative Reception at the Logical Layer

In the restoration of scientific meaning, the most serious threat is not silence but premature interpretation. Interpretation – especially when conducted without access to the original epistemic and experiential foundation – pollutes the logical structure of a concept and degrades its reliability. Therefore, a central tenet of regenerative reception is the **prohibition of interpretation** at the outset.

This does not mean suppressing meaning, but **delaying its construction** until the necessary components – structural integrity, primary records, and biological grounding – have been assembled. Meaning, in the context of scientific concepts, is not to be inferred from partial evidence or metaphorical association. It must emerge from a rigorous reconstruction based on the same logical architecture and empirical experience that generated the concept in the first place.

This principle is particularly relevant when dealing with contradictions detected by LLMs. These models are not to resolve the contradictions but to highlight them. Once identified, the contradictions must be examined without interpretation, to avoid importing assumptions or biases. Only after reassembling the logical premises – often through reference to first-person discovery accounts, laboratory notes, or theoretical justifications – can one proceed to restore the semantic content.

In this way, regenerative reception is not a matter of explaining a concept, but of re-living it. The scientific concept must be **regenerated in the consciousness of the learner**, with the same asymmetries, uncertainties, and structural necessities as in its original emergence. This requires, as previously argued, a biological substrate capable of storing not just the product of reasoning, but the reasoning process itself.

Ultimately, the prohibition of interpretation is an ethical safeguard. It protects the concept from distortion, the learner from false understanding, and the scientific community from cumulative error. Just as clinical medicine emphasizes differential diagnosis before treatment, regenerative reception insists on complete epistemological reconstruction before attribution of meaning.

6. Toward an AI-Assisted Framework for Contradiction Detection

To implement regenerative reception at scale, we must develop a systematic framework in which Large Language Models (LLMs) serve as partners in the

detection of contradictions, while remaining agnostic to interpretation. This chapter outlines such a framework and proposes criteria and procedures for its implementation.

Contradictions in scientific literature can take many forms: inconsistent definitions, conceptually incompatible metaphors, historical reinterpretations that obscure original intent, and terminological conflations. Many of these contradictions are imperceptible to human readers without extensive background knowledge, but they can often be identified by LLMs trained on vast textual corpora. The key is not to trust AI to resolve the contradiction, but to task it with exposing them reliably and reproducibly.

The proposed framework consists of three primary components:

Contradiction Typology and Tagging System: Contradictions are classified into categories – e.g., terminological, definitional, logical, translational – and tagged accordingly. This helps distinguish honest misinterpretation from epistemic error or structural omission.

Source Attribution and Priority Ranking: Each flagged contradiction is traced to its historical source. If primary records (e.g., notebooks, original publications, personal letters) are available, they are linked to the contradiction. In the absence of such documentation, the contradiction is tagged as low-priority or speculative, pending further archival discovery.

Human-Guided Regenerative Workflow: Human researchers are presented with flagged contradictions and their associated metadata. They are not expected to resolve the contradiction on the spot, but to initiate a process of regenerative reception – reconstructing the original logical structure and re-evaluating the meaning in light of both historical context and modern understanding.

This framework does not aim to accelerate scientific understanding in a linear way. Rather, it seeks to restore epistemic depth to concepts that have become flattened, diluted, or distorted by time. By combining algorithmic contradiction detection with biologically grounded human reasoning, we move toward a new paradigm of semantic integrity in interdisciplinary science.

In this context, LLMs are not interpreters but epistemic signalers. They illuminate where the historical thread has frayed. It is up to us to weave it back together.

7. Conclusion: Scientific Memory as a Regenerative Ecology

Scientific knowledge is not simply cumulative – it is recursive. Concepts do not progress in straight lines; they spiral through reinterpretation, contradiction, forgetting, and restoration. This paper has proposed a model of regenerative reception, wherein scientific concepts are not just transmitted or explained, but biologically and logically reconstituted in the consciousness of the learner.

We have shown that:

Scientific concepts differ fundamentally from everyday concepts in their logical structure and semantic memory basis.

B-cell networks in the cerebrospinal fluid may serve as a plausible biological substrate for concept integration.

Large Language Models can assist in contradiction detection but must remain confined to the physical layer.

Interpretation must be deferred until epistemological reconstruction has occurred.

A systematic framework for contradiction tagging, source attribution, and human-led reconstruction can support semantic integrity across disciplines.

Together, these principles outline a regenerative ecology of scientific memory – an environment where concepts are not just preserved but dynamically reactivated, where meaning is not received but **re-achieved**, and where machines serve not as surrogates for understanding but as tools that illuminate the paths we must walk again.

In a time when data multiplies faster than understanding, and language models simulate fluency without grounding, it is vital to reestablish the biological and logical conditions under which genuine knowledge emerges. Scientific memory, like biological memory, must not be archived – it must be **lived** anew.

REFERENCES

1. Piaget Jean. *The Psychology of Intelligence*. Routledge, 2001. [original: *La psychologie de l'intelligence*, 1947].
2. Vygotsky Lev S. *Thought and Language*. MIT Press, 1986. [original: *Myshlenieirech'*, 1934].
3. Jerne Niels K. “The Generative Grammar of the Immune System”. *Science* 229, no. 4718 (1985): 1057–1059.
4. Jerne Niels K. “Towards a Network Theory of the Immune System”. *Annalesd'immunologie* 125C (1974): 373–389.

Секция 3.

ЦИФРОВАЯ ТРАНСФОРМАЦИЯ ЯЗЫКОВОГО ОБРАЗОВАНИЯ

УДК 378

Аносова Наталия Эдуардовна

кандидат педагогических наук,
доцент Высшей школы лингвистики
и педагогики

Санкт-Петербургский политехни-
ческий университет Петра Великого
г. Санкт-Петербург, Россия

Natalia Anosova

PhD in Pedagogy, Associate Professor
of the Institute of Humanities

Peter the Great St. Petersburg

Polytechnic University

Saint-Petersburg, Russia

anosova_ne@spbstu.ru

Рубцова Анна Владимировна

доктор педагогических наук,
директор Высшей школы лингвистики
и педагогики, профессор Высшей

школы лингвистики и педагогики

Санкт-Петербургский политехни-
ческий университет Петра Великого
г. Санкт-Петербург, Россия

Anna Rubtsova

Doctor of Sciences in Education,

Head of the Graduate School

of Linguistics and Pedagogy,

Professor of the Institute of Humanities

Peter the Great St. Petersburg

Polytechnic University

Saint-Petersburg, Russia

rubtsova_av@spbstu.ru

ОБУЧЕНИЕ ПРОФЕССИОНАЛЬНО-ОРИЕНТИРОВАННОМУ
АНГЛИЙСКОМУ ЯЗЫКУ ЦИФРОВЫХ ЛИНГВИСТОВ
В УСЛОВИЯХ ЦИФРОВИЗАЦИИ

Целью данной статьи является изучение перспектив и проблем, связанных с обучением профессионально-ориентированному английскому языку специалистов в области цифровой лингвистики в условиях цифровизации и развития технологий искусственного интеллекта. Исследование было проведено в магистратуре по направлению «Цифровая лингвистика» в Санкт-Петербургском политехническом университете Петра Великого, на занятиях по английскому языку для профессионального общения. Методы исследования включают анализ учебного процесса и анкетирование студентов, обучающихся по направлению «Цифровая лингвистика». Авторы делают вывод о необходимости использования стратегий совместного обучения и цифровых технологий на занятиях по профессионально-ориентированному английскому языку для достижения более высокого уровня компетентности будущих специалистов и предлагают способы внедрения новых технологий в учебный процесс. В статье приведены фрагменты заданий в рамках разработанного учебного пособия.

Ключевые слова: цифровая лингвистика; цифровизация; совместное обучение; профессионально-ориентированный английский язык; магистратура; профессиональная подготовка.

TEACHING ENGLISH FOR SPECIFIC PURPOSES TO SPECIALISTS IN THE FIELD OF DIGITAL LINGUISTICS IN THE CONTEXT OF DIGITALIZATION

The article discusses the prospects and problems associated with teaching English for Specific Purposes to specialists in the field of digital linguistics in the context of digitalization. The study was conducted in the framework of Master's program in the field of "Digital Linguistics" at Peter the Great St. Petersburg Polytechnic University. The research methods included the analysis of the educational process and teaching aids and a survey of students studying in the field of "Digital Linguistics". The authors conclude that it is necessary to use a combination of various techniques of collaborative learning and digital technologies in the classes of professionally oriented English to achieve a high level of competence of future specialists. The study discusses the ways to introduce a variety of methods into the educational process. The article provides fragments of tasks within the framework of the developed teaching aid on professionally oriented foreign language.

Key words: digital linguistics; digitalization; collaborative learning; English for Specific Purposes; Master's degree; professional training.

Технологическое развитие общества характеризуется модернизацией экономической, геополитической и социально-культурной сфер жизни. Инновации в науке и технике находят отражение в разработке передовых методов преподавания иностранных языков. Процессы межкультурной интеграции требуют от выпускников университетов высокого уровня владения иностранным языком. В инженерных областях, в области информационных технологий и прикладных дисциплин, таких как цифровая лингвистика, существует острая необходимость в освоении профессионально-ориентированного английского языка (English for Specific Purposes) в связи с тем, что большое количество соответствующих материалов написано на английском языке, а профессиональное общение в основном осуществляется на английском языке. Профессиональная терминология, как и научная лексика, является основным компонентом обучения иностранному языку специалистов по инженерным направлениям и прикладным наукам. Владение иностранным языком на профессиональном уровне гарантирует трудоустройство выпускников и способствует выстраиванию успешной профессиональной траектории в определенной области знаний. Всё вышесказанное позволяет сделать вывод о необходимости постоянного совершенствования образовательного процесса в вузах и актуализации программ обучения иностранным языкам, что позволит вывести компетенции выпускников на новый уровень.

В настоящее время ключевым аспектом развития отечественной образовательной системы является цифровизация. Под влиянием процессов, связанных с цифровизацией, меняются не только бизнес-коммуникация, но и межличностные отношения людей, их образ и ритм жизни [1]. Цифровизация выводит социальное взаимодействие людей на качественно иной уровень, создавая новые модели научных, образовательных и других отношений [2]. Именно цифровизация во многом послужила толчком для появления

прикладных дисциплин, таких как цифровая лингвистика. Цифровая лингвистика представляет собой сравнительно новую область знаний, которая объединяет информационные технологии и лингвистику. Специалисты в этой области должны быть профессионалами в различных отраслях знаний, таких как автоматическая обработка текстов и анализ текстов с помощью компьютерных методов, функционирование систем машинного перевода и др. Хорошее владение английским языком крайне важно для специалистов именно в этой области, учитывая специфику этой области знания.

Достижение высокого уровня владения иностранным языком для профессионального общения всегда было предметом обсуждения педагогов и методистов [3]. Именно владение иностранным языком обеспечивает успешное общение с коллегами из других стран. Новые технологические процессы, новые социально-экономические условия предъявляют новые требования для преподавателей иностранных языков высшей школы. В условиях современного общества преподаватель иностранных языков в высшей школе должен быть хорошо знаком с процессами и сущностью цифровизации, должен уметь работать на ряде цифровых образовательных платформ и организовать учебный процесс на основе этих платформ. Цифровизация образования часто понимается как переход на электронную систему обучения на основе создания виртуальной информационной среды [4]. Соответственно, преподавателю высшей школы необходимо осваивать новые навыки в своей деятельности, которые относятся к информационному, технологическому и инструментальному компоненту компетенций преподавателя. Преподаватели ESP, работающие со студентами специальности «Цифровая лингвистика», должны быть готовы к постоянному совершенствованию, постоянному поиску ресурсов и разработке дидактических материалов и инструментов оценивания. Учитывая новые тенденции в высшем образовании, предоставляющие обучающимся возможность выстраивания индивидуальной образовательной траектории, преподавателю необходимо развивать индивидуальный подход к каждому студенту, учитывать его интересы и компетенции во владении иностранным языком, что, безусловно, сопряжено с применением вариативного набора обучающих методик. Важно разработать релевантные инструменты для общения преподавателя и студентов, позволяющие консультироваться друг с другом. В процессе обучения совместная деятельность, сотворчество преподавателя и студентов является необходимым условием эффективной профессиональной подготовки студентов. Немаловажным фактором оптимизации обучения выступает повышение познавательной активности студентов, чему в значительной степени способствует использование педагогической фасилитации [5].

С учетом современных тенденций была разработана программа дисциплины «Английский язык для профессионального общения». При разработке заданий в рамках учебной дисциплины были использованы стратегии совместного обучения, ментальные карты и интерактивные методы. В программу курса были включены следующие виды совместного обучения:

brain teaser – метод совместной деятельности, при котором один студент или группа студентов исследует определённую тему, а затем преподаёт её другим студентам; peer teaching – метод совместного обучения, при котором один студент обучает другого, причём первый выступает в роли эксперта, а второй – новичка; mind maps – метод совместного обучения, позволяющий наглядно проиллюстрировать связи между терминами или понятиями, рассматриваемыми в учебном материале; а также Socrates seminars – организованные студентами дискуссии на интересующие их профессиональные темы; case studies и др. Данные техники делают процесс обучения более эффективным и увлекательным. Для составления головоломок, кроссвордов, лексико-грамматических упражнений и тестов использовались инструменты искусственного интеллекта (ИИ), такие как Perplexity AI, Claude и другие нейронные сети. В ходе обучения также был выбран материал из открытого источника – конференции по практическому машинному обучению. Один из кейсов касался применения (ИИ) российскими IT-компаниями, такими как Сбер, Яндекс и другими. Студенты были распределены в малые группы, каждая из которых, на выбор, отвечала за презентацию одной из тем конференции: «GigaChat: наш опыт обучения LLM», «Компьютерное зрение роботов на складах Market», «Мультимодальные генеративные модели для обработки документов», «Подходы к централизованному управлению и мониторингу моделей машинного обучения в X5». Следует отметить, что подобные задания вызывают огромный интерес студентов и способствуют формированию не только профессиональных компетенций, но и так называемых soft skills. Примеры разработанных авторами заданий приведены ниже.

Задание 1. Разделитесь на группы и выберите кейс. Изучите кейсы, проанализируйте сценарии, подготовьте ответы на вопросы на английском языке (один ответ от каждого участника группы). Воспользуйтесь сетью Интернет для поиска дополнительной информации.

Case Study: NLP in Language Education Scenario: You are working as a linguist with a company developing a natural language processing (NLP) tool aimed at helping students learn new languages. This tool uses machine learning algorithms to analyze learners' mistakes in real-time and provide feedback on grammar, syntax, and vocabulary. However, teachers are concerned that the automated corrections might prevent students from developing deeper, intuitive language skills and are worried about the tool's accuracy when interpreting more complex or idiomatic language.

Questions: 1. What considerations should you take into account to ensure that the NLP tool supports rather than replaces traditional language learning methods? 2. How would you address concerns about the accuracy of the tool when it comes to complex language structures? 3. What steps would you recommend to balance automated feedback with teacher-guided learning?

Задание 2. Watch the video “ChatGPT raises questions about how humans acquire language Chatbots” and answer the questions below. Summarize the report in the source language and in the target language.

- *Chatbots are changing the Internet*
- *How do chatbots work?*
- *The problems with today’s chatbots*
- *The ELIZA effect*
- *Replika AI*
- *What might future chatbots be able to do?*
- *The drawbacks of chatbots*

Задание 3. Подготовьте устное сообщение на английском языке по одной из тем, представленных ниже. Выполните устный последовательный перевод на русский язык сообщений, подготовленных другими студентами.

- *Artificial intelligence can improve workplace safety.*
- *AI can offer accessibility for people with disabilities.*
- *AI can make everyday life more convenient and enjoyable, improving our health and standard of living.*
- *Artificial intelligence poses dangerous privacy risks.*
- *AI repeats and exacerbates human racism.*
- *AI will harm the standard of living for many people by causing mass unemployment as robots replace people.*

Специалисты в области цифровой лингвистики также должны обладать знанием математики и логики. Во время занятий мы внедрили технику совместного обучения (peer teaching), поощряющую студентов предлагать решения различных логических задач и делиться аргументами с остальными студентами. В качестве учебного материала были использованы логические материалы, пример одной из них приведен ниже.

Задание 4. Решите приведенную ниже задачу.

- *Of the 200 seniors at Monroe High School, exactly 40 are in the band, 60 are in the orchestra, and 10 are in both. How many students are in neither the band nor the orchestra? Choose the correct answer: A) 80; B) 90; C) 100; D) 110; E) 120.*

По окончании курса было проведено анкетирование студентов с целью выявления отношения студентов к различным видам заданий, представленных в разработанном курсе. Обучение проходило в течение трех семестров на 1–2 курсе магистратуры по направлению «Цифровая лингвистика». Большинство студентов (97 %) высоко оценили разнообразие учебных заданий, включенных в курс, отдельно отметив Socrates seminars и peer teaching как учебные техники, дающие возможность развить творческие навыки и умение работать в команде коллег-профессионалов. Обучение было организовано с использованием системы электронного обучения MOODLE, студентам был предоставлен свободный доступ к необходимым материалам и ресурсам, а также возможность прохождения тестирования в режиме онлайн. Подавляющее большинство студентов (89 %) высказались в пользу проведения

дискуссий и участия в заданиях с применением совместного обучения в режиме офлайн, что еще раз подтверждает исключительную ценность непосредственного общения преподавателя и студентов, в особенности, в условиях цифровизации образования. В рамках анкетирования были предложены следующие идеи: увеличение количества практических заданий, имитирующих те, которые выпускники будут выполнять в своей профессиональной деятельности; увеличение творческих заданий, стимулирующих критическое мышление; организация большего взаимодействия и непосредственного общения между преподавателями и студентами.

Следует отметить, что большинство студентов (87 %) осознают, что профессиональная конкурентоспособность во многом определяется компетенциями в области профессионально-ориентированного иностранного языка. Среди наиболее актуальных навыков были указаны умение обрабатывать большие объемы данных на иностранном языке, способность реализовать свои профессиональные функции в ситуациях, требующих знания иностранного языка, возможность участвовать в международных профессиональных проектах. Актуальность программ подготовки специалистов в области цифровой лингвистики сопряжена с постоянным совершенствованием учебных программ. Исследование показало, что дальнейшая разработка и совершенствование образовательных программ в области профессионально ориентированного языка является актуальной задачей для разработчиков в сфере образования.

ЛИТЕРАТУРА

1. Иванова О. В., Кутузова З. Ю. Иностранные языки в вузе: цифровые стратегии обучения // Ученые записки. Электронный научный журнал Курского государственного университета. 2023. №4 (68). URL: <https://cyberleninka.ru/article/n/inostrannye-yazyki-v-vuze-tsifrovye-strategii-obucheniya> (дата обращения: 15.06.2025).
2. Грязнова Е. В. Противоречия цифрового высшего образования в информационной культуре современного общества // Вестник Мининского университета. 2023. Т. 11. № 1. С. 12–20. DOI: 10.26795/2307-1281-2023-11-1-12.
3. Amanbay A. Development of professional communicative competence of foreign language teachers // Norwegian Journal of Development of the International Science, (97), pp.39–41 (2022).
4. Гордеева Е. В., Мурадян Ш. Г., Жажоян А. С. Цифровизация в образовании // Journal of Economy and Business. 2021. Т. 4–1 (74). С. 112–115.
5. Николаева Л. В., Саввинова Р. В. Взаимодействие преподавателя и студента как условие эффективности профессиональной подготовки будущих специалистов // Современные наукоемкие технологии. 2015. № 12–2. С. 351–354. URL: <http://www.top-technologies.ru/ru/article/view?id=35270> (дата обращения: 15.06.2025).

УДК 372.881.111.1

Быканова Надежда Николаевна
преподаватель кафедры
иностранных языков института общей
профессиональной подготовки
Национальный исследовательский
ядерный университет «МИФИ»
г. Москва, Россия

Nadezhda Bykanova
Lecturer of the Department of Foreign
Languages of the Institute of General
Professional Training
National Research Nuclear University
MEPhI
Moscow, Russia
bnn90@yandex.ru

ОБУЧЕНИЕ УМЕНИЯМ РЕФЛЕКСИИ И КОМПРЕССИИ ИНОЯЗЫЧНОГО ТЕКСТА В ПИСЬМЕННОЙ РЕЧИ СТУДЕНТОВ-ПРОГРАММИСТОВ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В современном мире искусственный интеллект оказывает все большее влияние на методы в обучении иностранным языкам, уже не выступая только лишь инструментом, к которому прибегают преподаватели. В данной статье рассматриваются некоторые цели обучения студентов-программистов иноязычной письменной речи, а также взаимосвязь между обучением письменной речи и чтению, при котором в основном и требуется умение компрессии текста, тогда как в письменной речи происходит декомпрессия внутренней речи в развернутую письменную форму. На основе этого предлагается комплекс упражнений, которые помогают развивать у студентов умения компрессии текста, предназначенные для работы в мини-группах с использованием запросов в чат-бот. Работа в мини-группах может оказывать положительное влияние на умение рефлексии и саморефлексии, которым также содействует обучение компрессии текста, а взаимодействие с чат-ботом позволяет не только разнообразить способы общения, но также рассматривать чат-бота как отдельного участника общения.

Ключевые слова: обучение письменной речи; искусственный интеллект; чат-бот; обучение рефлексии; обучение компрессии текста.

TEACHING REFLECTION AND FOREIGN LANGUAGE TEXT COMPRESSION SKILLS IN THE WRITTEN LANGUAGE USING ARTIFICIAL INTELLIGENCE

In the modern world, artificial intelligence is increasingly influencing the methods of teaching foreign languages, no longer acting as just a tool used by teachers. This article discusses some of the goals of teaching programming students to write in a foreign language, as well as the relationship between writing and reading, which mainly requires the ability to compress text, while writing decompresses internal speech into an expanded written form. Based on this, a set of exercises is proposed that help students develop text compression skills designed to work in mini-groups using chatbot queries. Working in mini-groups can have a positive impact on the ability of reflection and self-reflection, which is also facilitated by text compression training, and interaction with a chatbot allows not only to diversify communication methods, but also to consider a chatbot as a separate participant in communication.

Key words: teaching writing; artificial intelligence; chat-bot; teaching reflection; teaching text compression.

Обучение письменной речи предусматривается для всех ступеней образования и всех типов образовательных учреждений в рамках обучения письму как одному из четырех коммуникативных умений. Цели, задачи и методы обучения при этом разнятся для каждой ступени и типа учреждения. Для высших образовательных учреждений, подготавливающих специалистов в области информационных технологий и программирования, помимо глобальной цели в подготовке высококвалифицированных специалистов в области создания цифровых продуктов на уровне мировых стандартов [1, с. 3] ставится задача в воспитании кадров, обладающих сформированной иноязычной коммуникативной компетенцией и способных осуществлять деловую коммуникацию в устной и письменной формах на государственном и иностранном языках. Так как зачастую мотивом письменной речи являются мотивы основной деятельности, а не мотивы общения как таковые [2], логично будет предположить, что для выпускников вузов их непосредственная профессиональная деятельность и ее потребности определяют мотивы и те виды письменной речи, которым они должны овладеть.

Письменная речь определяется как структурная, целенаправленная система речевых действий в письменной форме, возникающая в процессе осуществления деятельности человека [3, с. 27]. М. А. Татаринова определяет иноязычную письменную речь как процесс формулирования и выражения мыслей для получения конечного результата, оформленного в виде текста [4, с. 13]. В данной статье под письменной речью мы будем понимать продуктивный вид речевой деятельности, характеризующийся структурной системой речевых действий и направленный на выражение собственных мыслей в письменной форме, возникающих в процессе осуществления деятельности человека, результат которой оформлен в виде текста. Письменный текст, по сравнению с разговорной речью, обладает рядом отличительных черт, таких как более стандартная форма лексических, грамматических и синтаксических средств, которые должны быть понятны его реципиенту, при этом пишущий рассматривает текст и с позиции читателя [2].

В широком понимании письменная речь это «форма речи, связанная с выражением и восприятием мыслей в графической форме» [5, с. 227] поэтому она объединяет продуктивную и рецептивную речевую деятельность, то есть письмо и чтение. Текст выступает как результат письменной речи и как коммуникативно-речевая единица при обучении чтению, что приводит к взаимосвязи обучения письменной речи и чтению. В рамках дисциплины в вузе происходит обучение всем четырем коммуникативным умениям: говорение, аудирование, чтение и письмо, однако в рамках коммуникативного подхода в последнее время закрепились тенденции к повышенному вниманию к обучению говорению и, принимая во внимание тот факт, что письмо и письменная речь долгое время рассматривались, в первую очередь, как средство обучения, а как его цель лишь отчасти, можно сделать вывод, что и обучение письменной речи, и обучение чтению страдают от недостатка внимания. Что касается мотивации самих студентов к чтению на иностранном языке и обучению (в том числе и

самообучению) ему, она возникает из профессиональных задач и соответствующих требований к кандидату на ту или иную должность, которые выдвигают работодатели на рынке труда и подчинена глобальной цели обучающегося стать высококвалифицированным и востребованным специалистом в своей области. Анализ существующих вакансий по запросу «программист» на hh.ru и требований к ним показал, что, несмотря на геополитическую ситуацию, владение английским языком как минимум на уровне, достаточном для чтения и понимания технических и нормативных текстов, как условие приема на работу, на сегодняшний день остается актуальным, что подчеркивает важность обучения умению чтения в вузе. Из этого можно сделать вывод, что мотивация студентов к овладению чтением достаточно высокая, что нельзя сказать о мотивации к овладению умением письменной речи, ввиду отсутствия видимого спроса среди работодателей и недостаточного внимания к проблеме со стороны преподавателей вузов.

Профессиональная деятельность и запросы рынка вакансий формируют перечень речевых произведений, которые будущий выпускник гипотетически может встретить после окончания вуза, а значит должен обладать необходимыми умениями их чтения и письма. К ним относятся: деловые и личные письма, а также ответы на них, которые предполагает деловая переписка, ответы на вопросы анкет и формуляров, резюме (CV), записки различного содержания, план, тезисы устного высказывания/письменного текста; рабочие записи, конспект; техническая документация на компьютерное программное обеспечение, пользовательская документация. Профессиональная направленность подобных текстов вкупе с повышенной мотивацией к чтению, уже имеющейся у студентов, позволяют утверждать о возможном использовании этих же речевых произведений для обучения письменной речи, а также наличия возможного положительного влияния на продуктивность ввиду взаимообусловленности деятельности и ее практической направленности.

Работа с текстом как с результатом собственной письменной речевой деятельности требует схожих умений что и чтение любых текстов, одним из которых является навык рефлексии и, если речь идет о собственном письменном высказывании, саморефлексии. Рефлексия лежит в основе понимания и заключается в выражении одного содержания в другом путем взаимных сопоставлений и противопоставлений. Развитию качества рефлексии способствует обучение компрессии текста [6, с. 73] для которой характерно выделение значимой информации, укрупнение элементов текста с сохранением логической структуры оригинального текста [7, с. 104]. По мнению Г. В. Ейгера и М. М. Гохлернера, обучение компрессии текста способствует развитию лингвистического мышления и ускорению выработки навыков беспереводного понимания речи на иностранном языке [8]. Беспереводное понимание, в свою очередь, помогает овладению письмом на иностранном языке, особенно на основе другого текста, так как он подразумевает его чтение, адекватное понимание и компрессию. Принимая во внимание, что процесс компрессии текста требует не только понимания общего смысла текста, но и анализа его содер-

жания, что может представлять трудности для студентов неязыкового вуза, целесообразно обучать компрессии текста с письменной речи, ввиду того что текст, как ее результат, рассчитан на зрительное восприятие с возможностью многократного прочтения написанного, что способствует его глубокому пониманию [9, с. 301; 10, с. 145].

Использование искусственного интеллекта (ИИ) и его эффективность в обучении рефлексии и саморефлексии студентов является предметом последних исследований ученых во всем мире. Так в исследовании, проведенном в 2024 г., студенты, получившие поддержку от чат-бота в заданиях на саморефлексию, показали повышение уверенности в знаниях и улучшили свои показатели [11]. Что касается обучению умениям компрессии текста с использованием ИИ, эта тема остается малоизученной. Так Е. В. Дзюба в своей статье о техниках искусственного интеллекта упоминает о возможном использовании ИИ для обучения компрессии текста на примере преподавания профессионально-ориентированного русского языка как иностранного [12]. Однако отдельного исследования о взаимосвязи навыков рефлексии, компрессии текста, а также последующей декомпрессии внутренней речи в развернутую письменную форму не проводилась.

Что же касается студентов-программистов, то искусственный интеллект представляет для них особый профессиональный интерес. Искусственный интеллект (ИИ) – это «комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их» [13]. К искусственному интеллекту относятся такие чат-боты как ChatGPT, YandexGPT, Perplexity, DeepSeek и т.д. Использование ИИ в образовании в последние годы только набирает обороты, но уже используется для научно-исследовательской деятельности (ГОСТ Р 70949-2023), научных публикаций (ГОСТ Р 71657-2024) и в образовании (ГОСТ Р 70945-2023), а также является предметом исследования многих отечественных ученых относительно области его применения в отношении индивидуализации обучения, оптимизации процесса подготовки преподавателя к занятиям и организация учебного процесса [14] в том числе качества оценивания письменной речи и обучения на основе обратной связи от ИИ. Многие исследователи отмечают эффективность использования ИИ в образовательном процессе [15, 16, 17]. Несмотря на это, искусственный интеллект и его применение в обучении иностранному языку еще мало изучены, в связи с этим отмечается необходимость в освещении проблемы его этичного использования и борьбы с антиплагиатом [18], а также обучении и педагогов и студентов искусству формулирования запросов к ИИ [14]. В связи с вышесказанным, логично предположить комплекс запросов к ИИ и упражнений на развитие умений компрессии текста и декомпрессии внутренней речи в развернутую письменную форму, которые будут задействовать возможности искусственного интеллекта.

Так, до начала работы с ИИ преподавателю необходимо рассказать студентам об этике его использования, а также о возможности дезинформации в ответах на запросы, возникающей в силу технического развития искусственного интеллекта, который может «галлюцинировать» – выдавать придуманные факты за реальность, ввиду недостаточной доступной информации. После ознакомления с базовыми правилами использования ИИ и чат-ботов, студентам в качестве домашней подготовки к аудиторному занятию может быть предложен к прочтению профессиональный текст (статья, инструкция или другая техническая документация) с последующим составлением его тезисного пересказа или краткого содержания. На аудиторном занятии, в результате групповой работы (предпочтительно в мини-группах по 3–4 человека), на основе заранее подготовленных студентами тезисных пересказов, предлагается обсудить общий на группу вариант пересказа. После этого, группе дается задание создать запрос в чат-бот, например: *«Проанализируй следующий текст и сделай его тезисный пересказ на английском языке, который будет отображать все важные элементы содержания, сохранять структуру повествования, обобщать основные идеи и давать отсылки на важные детали такие как имена, годы, цифровые данные»*. После этого в группах студенты должны проанализировать и сравнить ответ, предоставленный чат-ботом и свой тезисный пересказ, а также дать качественную оценку обоим и на основе данного анализа устранить существующие недочеты в собственном тезисном пересказе, если таковые были найдены. Это позволит провести рефлексию навыков студентов к компрессии текста. Для развития умений декомпрессии внутренней речи в развернутую письменную форму, данное задание может быть видоизменено и дополнено. Так, на этапе домашней подготовки к аудиторному занятию все студенты делятся на две большие группы, где каждой из них даются разные по содержанию, но одинаковые по своей структуре тексты. Во время аудиторного занятия студенты в мини-группах работают по алгоритму со своими текстами и их тезисными пересказами. После этого группы обмениваются своими тезисными пересказами, по которым им нужно будет составить развернутое письменное высказывание в стиле того текста, с которым они уже работали. Так как письмо требует достаточно много времени, написание развернутого письменного высказывания на основе тезисного пересказа можно оставить в качестве домашнего задания. По результатам работы на следующем занятии предлагается рефлексия в тех же самых мини-группах, между которыми происходил обмен, сравнение результатов проделанной самостоятельной работы, а также запрос в чат-бот о создании развернутого высказывания на основе предлагаемого тезисного пересказа. Для этого в чат загружается сам тезисный пересказ, а также текст похожего стиля, с которым студенты работали ранее, но отличающийся по содержанию от того, по которому был написан тезисный пересказ и отправляется запрос: *«Создай развернутое письменное высказывание на основе его тезисного пересказа. Для примера используй прикрепленный текст»*. После получения ответа от чат-бота, в мини-группах сравниваются уже варианты развернутого письменного высказывания, пре-

доставленные студентами и ИИ, производится их анализ. В конце всем студентам становятся доступны оба оригинальных текста и тем, кто уже работал с оригинальным текстом предлагается оценить все три полученных варианта развернутого письменного высказывания в контексте сходства с оригинальным текстом, полноты раскрытия информации, а также дать обратную связь.

Использование чат-ботов вкупе с работой в группах позволяет студентам развивать самостоятельность в организации своей работы. Предоставляемая обратная связь от одноклассников и ИИ позволяет развивать умение рефлексии и саморефлексии. Обучение навыку компрессии текста, а также декомпрессии внутренней речи в развернутую письменную форму может позволить снизить когнитивную нагрузку благодаря особенностям письменной речи по сравнению с устной и, в связи с этим, добиться более положительных результатов. Использование ИИ, представляющего профессиональный интерес для студентов-программистов, создает ситуацию повышенного интереса для студента, а также иллюзию свободы действия и выбора, без необходимости быть оцененным и проверенным преподавателем, что может повысить мотивацию к обучению письменной речи. Стоит отметить, что качественный и количественный анализ влияния использования чат-ботов на качество обучения умению компрессии и декомпрессии при обучении письменной речи может стать предметом следующих исследований.

ЛИТЕРАТУРА

1. Характеристика образовательной программы. Компетентностная модель выпускника: математическое и программное обеспечение вычислительных машин и компьютерных сетей : образовательная программа 09.03.04 "Программная инженерия" : уровень образования – бакалавриат / Национальный исследовательский ядерный университет "МИФИ", Институт интеллектуальных кибернетических систем ; утв. и.о. проректора Весна Е. Б. Москва : НИЯУ МИФИ, 2023. № 385 в реестре образовательных программ.
2. Эльконин Д. Б. Развитие устной и письменной речи учащихся / под ред. В. В. Давыдова, Т. А. Нежновой. М. : ИНТОР, 1998. 112 с. URL: <https://psychlib.ru/mgppu/Eru-001/Eru-001.htm> (дата обращения: 15.06.2025).
3. Леонтьев А. А. Язык, речь, речевая деятельность. М. : Просвещение, 1969. 214 с.
4. Татарина М. А. Теоретические основы создания и использования дистанционного курса обучения иноязычной письменной речи для студентов 2–3 курсов лингвистического вуза : На материале английского языка : автореф. дис. ... канд. пед. наук : 13.00.02. Москва, 2005. 28 л.
5. Азимов Э. Г., Щукин А. Н. Новый словарь методических терминов и понятий (теория и практика обучения языкам). М. : Изд-во ИКАР, 2009. 448 с.
6. Богин, Г. Филологическая герменевтика. Калинин : Калининский гос. ун-т, 1982. 48 с.
7. Сорокина И. Г., Цветкова Т. К. Обучение письму как компонент подготовки по иностранному языку // Вопросы филологии. 2010. № 1(34). С. 102–113.

8. Ейгер Г. В., Гохлернер М. М. О роли овладения приемами компрессии речи при обучении неродному языку // Психология грамматики / Под ред. А. А. Леонтьева, Т. В. Рябовой. М., 1968. С. 30–37.
9. Щукин А. Н. Методика обучения речевому общению на иностранном языке. Учебное пособие для преподавателей и студентов языковых вузов. М. : Издательство Икар, 2017. 454 с.
10. Прокофьева О. А. Формирование навыков письменной речи // Университетские чтения. 2019. С. 145–149.
11. Saritepeci M., Yildiz Durak H. Effectiveness of artificial intelligence integration in design-based learning on design thinking mindset, creative and reflective thinking skills: An experimental study [Electronic resource] // Educ Inf Technol 29, 25175–25209 (2024). DOI: <https://doi.org/10.1007/s10639-024-12829-2>.
12. Дзюба Е. В. Уроки русского для иностранцев: техники искусственного интеллекта или искусство техники? // Русистика. 2024. Т. 22. №2. С. 242–261. DOI: 10.22363/2618-8163-2024-22-2-242-261.
13. О развитии искусственного интеллекта в Российской Федерации: Указ Президента Российской Федерации от 10.10.2019 № 490 (в ред. № 124 от 15.02.2024). URL: <http://www.kremlin.ru/acts/bank/44731> (дата обращения: 28.10.2024).
14. Сысоев П. В. Искусственный интеллект в образовании: осведомлённость, готовность и практика применения преподавателями высшей школы технологий искусственного интеллекта в профессиональной деятельности // Высшее образование в России. 2023. №10. URL: <https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-obrazovanii-osvedomlyonnost-gotovnost-i-praktika-primeniya-prepodavatelyami-vysshey-shkoly-tehnologiy> (дата обращения: 06.05.2025).
15. Четырина Н. В. Организационно-педагогические условия обучения студентов иноязычному письменному взаимодействию на основе практики с чат-ботами // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2024. Т. 29. № 6. С. 1590–1607. DOI 10.20310/1810-0201-2024-29-6-1590-1607.
16. Попович О. Ю. Эффективность использования YandexGPT для развития речи по русскому языку как иностранному // Мир науки. Педагогика и психология. 2024. Т. 12. № 4. URL: <https://mir-nauki.com/PDF/31PDMN424.pdf> (дата обращения: 23.04.2025).
17. Преподаватель vs искусственный интеллект: сравнение качества предоставляемой преподавателем и генеративным искусственным интеллектом обратной связи при оценке письменных творческих работ студентов / Сысоев П. В., Филатов Е. М., Хмаренко Н. И., Мурунов С. С. // Перспективы науки и образования. 2024. № 5 (71). С. 694–712. DOI: 10.32744/pse.2024.5.41.
18. Сысоев П. В. Этика и ИИ-плагиат в академической среде: понимание студентами вопросов соблюдения авторской этики и проблемы плагиата в процессе взаимодействия с генеративным искусственным интеллектом // Высшее образование в России. 2024. Т. 33. № 2. С. 31–53. DOI: 10.31992/0869-3617-2024-33-2-31-53.

УДК 37.014.3:004.8:81

Головач Елена Игоревна

кандидат педагогических наук,
доцент, доцент кафедры фонетики
и грамматики английского языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Yelena Golovatch

PhD in Pedagogy, associate professor,
associate professor of the department
of English grammar and phonetics
Belarusian State University
of Foreign Languages
Minsk, Belarus
golovatch@yahoo.com

Кохан Маргарита Николаевна

старший преподаватель кафедры
интенсивного обучения иностранным
языкам
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Margarita Kochan

Senior Lecturer of the FL intensive
teaching department
Belarusian State University
of Foreign Languages
Minsk, Belarus
margaritakohan@gmail.com

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБУЧЕНИИ ИНОСТРАННОМУ ЯЗЫКУ

В статье рассматривается использование искусственного интеллекта (ИИ) в обучении иностранным языкам. Описываются основные технологии ИИ для создания текстового, визуального, аудио- и видеоконтента, а также интерактивных упражнений и презентаций. Особое внимание уделяется таким инструментам, как ChatGPT, Perplexity, Curipod, ElevenLabs, Character.ai и MagicSchool, которые помогают автоматизировать проверку заданий, создавать учебные материалы и развивать умения устной и письменной речи. Отмечается важность правильного формирования запросов к ИИ для получения качественных результатов. Также обсуждаются ограничения и риски ИИ: возможность генерации недостоверной информации, проблемы с авторскими правами и отсутствие критического мышления у нейросетей. Авторы подчеркивают необходимость осознанного и критического использования ИИ, чтобы сохранить творческий подход и самостоятельность обучающихся и преподавателей.

Ключевые слова: искусственный интеллект; нейросети; образовательные платформы; промпты; контент.

USING ARTIFICIAL INTELLIGENCE IN TEACHING A FOREIGN LANGUAGE

The article discusses the use of artificial intelligence (AI) in foreign language learning. It describes key AI technologies for creating text, visual, audio, and video content, as well as interactive exercises and presentations. Special attention is paid to tools such as ChatGPT, Perplexity, Curipod, ElevenLabs, Character.ai, and MagicSchool, which help automate assignment checking, create educational materials, and develop oral and written language skills. The importance of properly formulating AI prompts to achieve high-quality results is emphasized. The article also addresses the limitations and risks of AI, including the possibility of

generating inaccurate information, copyright issues, and the lack of critical thinking in neural networks. The authors stress the need for conscious and critical use of AI to preserve creativity and independence for both learners and teachers.

Key words: artificial intelligence; neural networks; educational platforms; prompts; content.

В последнее время в практике преподавания иностранных языков все шире используются информационно-коммуникационные технологии. Особое место среди них занимает искусственный интеллект (ИИ). С его помощью преподаватель и обучающиеся могут генерировать контент разных типов: текстовый, визуальный, аудио и видео контент, а также сочетание контента разных типов. В зависимости от типа генерируемого контента можно выделить следующие технологии искусственного интеллекта:

- технологии ИИ для создания текстов, анализа текстов, генерации речи (*Chat GPT, Perplexity, Wordvice, Gamma*, etc);
- технологии ИИ для создания презентаций (*curipod.com*);
- технологии ИИ для создания изображений (*shedevrum*, etc);
- технологии ИИ для создания озвучки текстов и создания видео (*ElevenLabs, naturalreaders.com*, etc);
- технологии ИИ для создания упражнений, заданий, специально для обучения иностранным языкам (*Twee, MagicSchool, Adcafe* etc);
- интерактивные чаты для развития умений говорения или письменной речи (*talkpal, Character.ai, gliglish* etc).

Одними из наиболее часто используемых (универсальных) технологий искусственного интеллекта являются *Chat GPT* и *Perplexity*. С их помощью преподаватель может решать целый спектр профессиональных задач: от проверки письменных заданий до создания комплексов упражнений. Для успешной работы с данными технологиями необходимо грамотно формулировать промпты, т.е. инструкции, задающие задание искусственному интеллекту. Промпты необходимо формулировать максимально конкретно и подробно, указывать тему, типы упражнений, например, *a matching exercise, a gap filling exercise, questions for discussion, communicative situations for role-plays*. Если упражнения нацелены на формирование и совершенствование лексических навыков по определенной теме, необходимо списком указывать лексические единицы, которые должны быть использованы в упражнениях. Еще одним важным условием успеха является правильно указанный уровень сложности, например, A1, B1 и так далее. Если уровень не указан, то искусственный интеллект, как правило, создает упражнения на уровень B2.

С помощью ИИ можно проверять письменные задания. Для этих целей используется как универсальные инструменты типа *Chat GPT* и *Perplexity*, так и специально созданные для этих целей технологии, например, *Wordvice* и *Gamma*. С их помощью можно не только проверить тексты на наличие лексических и грамматических ошибок, но и исправить стилистические ошибки, например, неверный выбор регистра. Искусственный интеллект может только выделять и корректировать ошибки, исправления можно удалить и предложить обучающимся самим исправить подчеркнутые ошибки.

Искусственный интеллект может помочь в генерации увлекательных презентаций, обеспечивая активное участие и вовлеченность обучающихся. Зарегистрировавшись на сайте *curipod.com*, преподаватель имеет возможность сам создать презентацию, включив туда такие интерактивные слайды, как опрос участников, облако слов, открытые вопросы, вопросы с множественным выбором, задания по рисованию и визуализации, обратную связь, предлагаемую искусственным интеллектом. На сайте есть кнопка предпросмотра презентации, чтобы преподаватель мог увидеть, как будет выглядеть презентация у него на экране и на экранах обучающихся. Преподаватель показывает готовую презентацию на мультимедийной доске, а обучающиеся получают к ней доступ по ссылке, пин-коду или qr-коду, используя свои мобильные телефоны, компьютеры или планшеты для ответов на вопросы. Все ответы обучающихся можно увидеть на экране и прокомментировать. Если у преподавателя уже есть готовая презентация, которую он хотел бы сделать более динамичной, то ее можно загрузить на сайт, указать возраст или уровень обучающихся, выбрать язык, при желании добавить дополнительные критерии, и система сгенерирует свою презентацию, добавив в исходную интерактивные слайды и элементы, адаптированные к ее теме.

Система может и сама сгенерировать урок по запросу преподавателя. Нужно указать предмет обучения, язык, задачу и цели урока (презентация языкового явления, формирование или совершенствование навыков, развитие умений, повторение пройденного и т.д.) и искусственный интеллект предложит свою презентацию, состоящую из 10–15 слайдов, включая интерактивные инструменты. Преподаватель может вносить в нее изменения, добавлять или удалять слайды. Данный инструмент значительно упрощает процесс создания презентаций и повышает заинтересованность и мотивацию обучающихся.

Если необходимо создать изображение, можно использовать одну из многочисленных технологий генерации визуального контента, например, *shedevrum.ai*. С помощью этой нейросети можно создавать изображения и в разных стилях: рисованном карандашом, красками, в стиле фотографии. Можно «оживлять картинки», создавая короткие видеоролики.

Кроме того, нейросети помогают преподавателям создавать кастомизированные аудио и видеоматериалы. На сайтах *elevenlabs.io* и *naturalreaders.com* можно озвучить тексты, выбрав мужской или женский голос, нужный акцент, тембр голоса и скорость, причем *naturalreaders.com* предлагает также опцию озвучивания диалогов разными голосами, а на *elevenlabs.io* есть возможность добавить звуковые эффекты, например, скрип тормозов или шум дождя.

Для создания видео предлагаем воспользоваться сайтом *invideo.io*. Преподаватель может быстро превратить свои тексты или статьи в видеоролики, выбрав нужные шаблоны. Редактор разбивает текст на сцены и добавляет видеоклипы и изображения из обширной библиотеки. Нейросеть способна и сама создать видеоролик по запросу преподавателя, который указывает тему

и описывает свою идею, по желанию добавляет список слов и выражений, которые должны быть включены в видео, определяет длительность видеоролика (от 1 минуты до 10 минут), указывает аудиторию и платформу (*Youtube, Instagram, TikTok*) и в течение 10–15 минут получает сгенерированный видеоролик. Если преподавателя не совсем устраивает результат, то можно с помощью подсказок попросить нейросеть внести изменения, например, сделать видео более коротким или более смешным.

Для развития умений устной речи можно использовать чат-боты с искусственным интеллектом. Например, на сайте *character.ai* обучающиеся могут выбрать для общения любого персонажа: любимого героя фильма, актера, известного бизнесмена, фитнес тренера, профессионального психолога или учителя английского языка. При желании они могут создать свой персонаж для общения. Персонаж сам начнет беседу, написав сообщение. Обучающиеся могут написать ответ или прослушать сообщение и «позвонить» собеседнику, разрешив программе использовать микрофон. Сайт *gliglish.com* предоставляет возможность общения с ИИ в двух режимах: можно разыграть диалог по выбранному сценарию или поговорить на свободную тему, при этом программа будет предлагать наводящие вопросы и даже подсказки, которые помогут начать ответ. Ответ записывается и выводится на экран с исправленными ошибками.

В мессенджере *Telegram* существует специальный бот для изучения английского языка *Papaya English Learning AI Character*. Бот попросит обучающихся указать свой уровень владения языком, спросит, для чего они хотят изучать язык, а также предложит выбрать несколько интересующих их разговорных тем. На эти темы с ним можно общаться путем обмена письменными или голосовыми сообщениями. Бот обещает оценить уровень владения языком после 5 сообщений. Процесс общения происходит очень реалистично, и при этом обучающиеся не будут испытывать дискомфорт, неуверенность и страх не понять вопроса или быть неправильно понятыми, как это иногда случается при разговоре с англоязычными собеседниками. Искусственный интеллект будет повторять вопросы или переформулировать их столько раз, сколько потребуется. Обучающиеся смогут узнать множество новых слов и выражений по определенной теме, которые будут использованы в сообщениях программы.

На сайте *smalltalk2.me* есть возможность протестировать свой уровень владения языком, разговаривая с искусственным интеллектом. В бесплатной версии предлагается одна попытка, чтобы ответить на ряд вопросов, сгенерированных программой. Это занимает около 12–15 минут. После завершения теста обучающиеся получают подробный анализ своей речи с комментариями и рекомендациями.

Образовательные платформы с искусственным интеллектом предлагают преподавателю обширный набор инструментов для создания учебных материалов. Например, на платформе *magicschool.ai* таких инструментов больше 60. Рассмотрим некоторые из них, которые можно успешно применять

для обучения английскому языку. Существует функция, который адаптирует текст под уровень владения языком, позволяя преподавателю выбрать любую аутентичную статью и адаптировать ее для работы со своей аудиторией. Нейросеть также создает тесты с вопросами множественного выбора или открытыми вопросами по загруженному тексту для того, чтобы проверить понимание его содержания. Для развития умений аудирования нейросеть может создать вопросы по видеоролику из *Youtube*. Для работы с вокабуляром нейросеть может сгенерировать текст разного объема, от одного абзаца до нескольких страниц, включив туда указанные преподавателями слова и выражения. Для того, чтобы сделать занятие более увлекательным и занимательным, преподаватель имеет возможность воспользоваться инструментами генерации скороговорок и шуток на изучаемую тему. Их предлагается использовать на этапе речевой разминки или в конце занятия. Для совершенствования навыков говорения преподаватель может также воспользоваться функцией генерации начала предложений. Указав тему, преподаватель получает несколько фраз, которые помогут обучающимся начать свое высказывание. Нейросеть также поможет в планировании урока, описав начало урока, предложив виды упражнений и режимы работы, а также домашнее задание.

Twee.com – это платформа, специально созданная для преподавания иностранных языков. С ее помощью можно создавать упражнения для текстов для чтения и видео (по ссылке с *youtube*), упражнения на формирование и совершенствование лексических навыков, например, на соотнесение слов и выражений с дефинициями, заполнение пропущенных слов в предложениях. На платформе есть возможность создания коммуникативных заданий, например. Вопросы для обсуждения по определенной теме, а также заданий на *warm-up*. Кроме того, с помощью *twee.com* можно делать скрипты видео, а также создавать тексты разных жанров (письмо, учебных текст, эссе, диалоги) на различные темы с использованием определенной лексики. Нейросеть даже генерирует планы занятий. Единственный минус данной образовательной платформы заключается в том, что после короткого испытательного периода ее использование становится платным, а самые востребованные инструменты вообще нельзя использовать бесплатно.

Приведенный в данной статье перечень технологий искусственного интеллекта, используемых в практике преподавания иностранных языков, ни в коей мере не является полным, возможности ИИ постоянно расширяются, возникают новые платформы и нейросети, которые помогают сэкономить время и усилия преподавателей для подготовки к занятиям и проверки письменных работ.

При всех очевидных преимуществах использования ИИ в практике преподавания не стоит забывать и о недостатках, ловушках, в которые легко можно попасть. Корни проблем, как правило, проистекают из ошибок в исходном датасете, источниках, используемых ИИ, некорректных запросах со стороны пользователей и, наконец, в отсутствии у ИИ настоящего понимания сути вещей и критического мышления.

Одной из наиболее опасных ловушек является явная ложь, галлюцинирование ИИ, особенно ярко это может проявиться при попытках использования ИИ для написания курсовых, дипломных работ или научных статей. Можно столкнуться с тем, что нейросети ссылаются на несуществующие источники, объединяют несколько источников в один, ссылаются на несуществующих ученых. Так, например, при написании курсовой работы, посвященной взаимосвязанному обучению аудированию и говорению, ИИ сослался на Б. А. Голицына как авторитетного ученого-психолингвиста. Такой человек действительно существовал, но был боярином при дворе царя Алексея Михайловича.

Еще одной проблемой использования нейросетей являются неясности с авторскими правами. Законодательно использование ИИ на данный момент никак не закреплено. И, наконец, постоянное использование нейросетей не позволяет развивать критическое мышление и самостоятельный поиск ответов на вопросы, креативность. Это утверждение, по нашему мнению, справедливо, как для студентов, так и для преподавателей.

УДК 37.01

Захарова Елена Валерьевна

кандидат филологических наук,
доцент кафедры русского языка,
литературы и журналистики
ФГБОУ ВО «УлГПУ
им. И. Н. Ульянова»
г. Ульяновск, Россия

Elena Zakharova

PhD in Philology, Associate Professor
of the Department of Russian
Language, Literature and Journalism
Ulyanovsk State Pedagogical
University named after I. N. Ulyanov
Ulyanovsk, Russia
zaharova2007@yandex.ru

Куварина Екатерина Николаевна

старший преподаватель кафедры
иностранных языков
ФГБОУ ВО «УлГПУ
им. И. Н. Ульянова»
г. Ульяновск, Россия

Ekaterina Kuvarina

Senior Lecturer of the Department
of Foreign Languages
Ulyanovsk State Pedagogical
University named after I. N. Ulyanov
Ulyanovsk, Russia
en.kuvarina@gmail.com

ИНТЕГРАЦИЯ СРЕДСТВ ДОПОЛНЕННОЙ РЕАЛЬНОСТИ В МЕТОДИКУ ПРЕПОДАВАНИЯ РКИ: МОДЕЛИ И ПРАКТИКИ

В данной статье рассматривается проблема интеграции технологии дополненной реальности (AR) в методику преподавания русского языка как иностранного (РКИ) с целью формирования социокультурной компетенции обучающихся. Проанализированы зарубежные и отечественные подходы к использованию AR в образовательном процессе, выявлены их особенности. Проведен сравнительный анализ моделей SAMR, TRACK и FRAME с точки зрения их применимости к обучению РКИ на начальном уровне (A1). Приведены результаты опроса педагогов. В качестве иллюстраций рассмотрены примеры заданий по теме "Русская кухня" из учебно-методического пособия для изучающих русский язык как иностранный "Приглашение в Россию". Дальнейшие исследования могут включать в себя попытки интегрировать все три модели с целью расширения методической базы, адаптацию под конкретные запросы при изучении русского языка как иностранного, что позволит создать более персонализированные образовательные траектории и сценарии для изучающих РКИ.

Ключевые слова: русский язык как иностранный; модели; методика преподавания РКИ.

Настоящая работа выполнена в рамках проекта «Формирование социокультурной компетенции при обучении РКИ в Федеративной Демократической Республике Эфиопии на основе технологии дополненной реальности» государственного задания Министерства просвещения РФ № 073-03-2025-066/2 от 25.04.2025 г.

INTEGRATION OF AR TOOLS INTO THE METHODOLOGY OF TEACHING RUSSIAN AS A FOREIGN LANGUAGE: MODELS AND PRACTICES

This article examines the problem of integrating augmented reality (AR) technology into the methodology of teaching Russian as a foreign language (RFL) in order to form the socio-

cultural competence of students. The article analyses foreign and domestic approaches to the use of AR in the educational process, and identifies their features. A comparative analysis of the SAMR, TPACK and FRAME models has been carried out in terms of their applicability to RFL training at the elementary level (A1). The results of a survey of teachers are presented. Russian cuisine examples from the "Invitation to Russia" textbook for students of Russian as a foreign language are considered as illustrations. Russian studies may include attempts to integrate all three models in order to expand the methodological base and adapt them to specific needs when learning Russian as a foreign language, which will create more personalized educational trajectories and scenarios for learners of RFL.

Key words: Russian as a foreign language; models; teaching methodology of Russian as a foreign language.

В последние годы наблюдается рост интереса к использованию технологии дополненной реальности в образовательном процессе. Традиционные методы и приемы обучения замещаются методами, основанными на современных технологиях, отличающихся от традиционных приемов интерактивностью, погружением в обучающую среду и персонализацией. Ряд зарубежных и отечественных исследований доказывает положительный эффект AR-технологии на рост вовлеченности, мотивации, улучшение запоминания новой лексики и грамматических структур за счет наглядности, возникновения визуальных ассоциаций, контекстуализации и активности обучающихся.

Так, например, Майер и Фиорелла отмечают, что одновременное использование аудиального и визуального восприятия в сочетании с активной когнитивной деятельностью ускоряет и повышает качество усвоения материала [1]. В отечественной науке также заметно увеличился интерес к технологии дополненной реальности. Так, отмечается рост эффективности обучения, снятие психологических барьеров, появляются исследования, затрагивающие проблемы РКИ [2, 3].

Однако подходы к применению технологий дополненной реальности в процессе обучения в российской и зарубежной науке существенно различаются, что делает актуальным проведение их анализа (Таблица 1).

Таблица 1

Сравнительный анализ подходов к использованию технологий дополненной реальности в России и за рубежом

Критерий	Отечественная педагогика	Зарубежный подход
основные цели	Повышение наглядности, мотивации; создание аутентичных ситуаций общения	Индивидуализация траектории обучения; активное вовлечение; создание иммерсивной учебной среды
методика использования	Интеграция в традиционные занятия; акцент на визуализацию и моделирование ситуаций, приближенных к реальным коммуникативным ситуациям	Применение моделей SAMR, TPACK; адаптивные сценарии

уровень внедрения	Часто ограниченный, экспериментальный; пилотные проекты	Более широкое внедрение в систему образования на ступени школа и вуза; наличие методических рекомендаций
компетенции	Фокус на языковой и социальной компетенции	Фокус на когнитивные, метапознавательные и коммуникативные навыки; развитие критического мышления
ограничения	Ограниченность оснащения и высокая стоимость; отсутствие методической базы и подготовки педагогов	Ограниченность оснащения; необходимость технической поддержки

Проведенный сравнительный анализ демонстрирует как сходства, так и различия между подходами к технологии дополненной реальности в России и за рубежом. В целом, в зарубежной педагогической практике можно отметить более системный подход, характеризующейся наличием разнообразных моделей и наличием методических указаний. Такое положение можно объяснить тем, что в России AR-технологии менее изучены и только начинают внедряться преимущественно в виде экспериментальных и пилотных проектов, интеграции в традиционные занятия. Можно отметить отсутствие методических моделей, рекомендаций, систем.

Однако, одной из ключевых проблем является отсутствие единой методической модели использования AR-технологии в языковом обучении как в отечественной науке, так и за рубежом. В англоязычной литературе в контексте изучения данной технологии упоминается использование различных моделей, ориентированных на преподавание английского языка и требующие адаптации к специфике РКИ из-за необходимости учета социокультурных аспектов, связанных с изучением русского языка.

Исходя из этого появляется необходимость анализа моделей, применяемых за рубежом с целью адаптации наиболее подходящего варианта к эксперименту, реализуемому на базе Ульяновского государственного педагогического университета имени И. Н. Ульянова «Формирование социокультурной компетенции при обучении РКИ в Федеративной Демократической Республике Эфиопии на основе технологии дополненной реальности» в рамках государственного задания Министерства просвещения РФ № 073-03-2025-066/2 от 25.04.2025 г.

В научной литературе активно обсуждаются и используются различные модели для описания процесса внедрения и применения различных цифровых технологий в образование. Среди них наибольшей популярностью пользуются следующие модели: SAMR, TPACK и FRAME [4; 5; 6; 7]. Проведенный нами анализ можно кратко представить в Таблице 2.

Таблица 2

Сравнительный анализ моделей SAMR, TPACK и FRAME для работы с AR
в процессе обучения русскому языку как иностранному

Критерий	SAMR (Substitution, Augmentation, Modification, Redefinition)	TPACK (Technological pedagogical content knowledge)	FRAME (Framework for the Rational Analysis of Mobile Education)
основная идея	Иерархия уровней использования технологий: замещение, расширение (дополнение), модификация и преобразование	Взаимосвязь педагогических, технологических и предметных знаний педагога	Мобильное обучение через взаимосвязь мобильное устройство с AR – ученик – социальная среда
сильные стороны в рамках нашей задачи	Простота в структуре, позволяет адекватно и легко оценить глубину внедрения технологий. Носит прикладной характер	Позволяет регулировать самому педагогу количество использований технологий в процессе обучения – педагогическое проектирование	Принимает во внимание социальный контекст
слабые стороны	Недостаточно проработана методическая составляющая	Отсутствуют уровни интеграции, в связи с чем возрастает нагрузка на педагога	Недостаточно учитывает специфику языкового обучения
потенциал использования модели применительно к AR	Высокий: удобно оценивать на каком из уровней используется технология дополненной реальности	Средний: помогает понять, какие шаги нужны для применения технологии, но не имеет шкалы глубины внедрения	Высокий: учитывает специфику мобильных устройств, но нет методики внедрения технологии
потенциал использования в конкретной ситуации (уровень знания русского языка А1, ограниченное время)	Оптимальный: можно реализовать эффективные и простые сценарии уже на первых двух уровнях	Средний: полезен в проектировании	Средний: помогает учесть все факторы, но необходимо больший временной промежуток

Проведенный опрос среди педагогов УлГПУ на сравнение данных моделей по 4 критериям: наглядность, необходимость подготовки педагога, учет особенностей мобильных устройств или контекста, соответствие технологии уровню владения языком позволил визуально представить полученные данные (Рисунок 1).

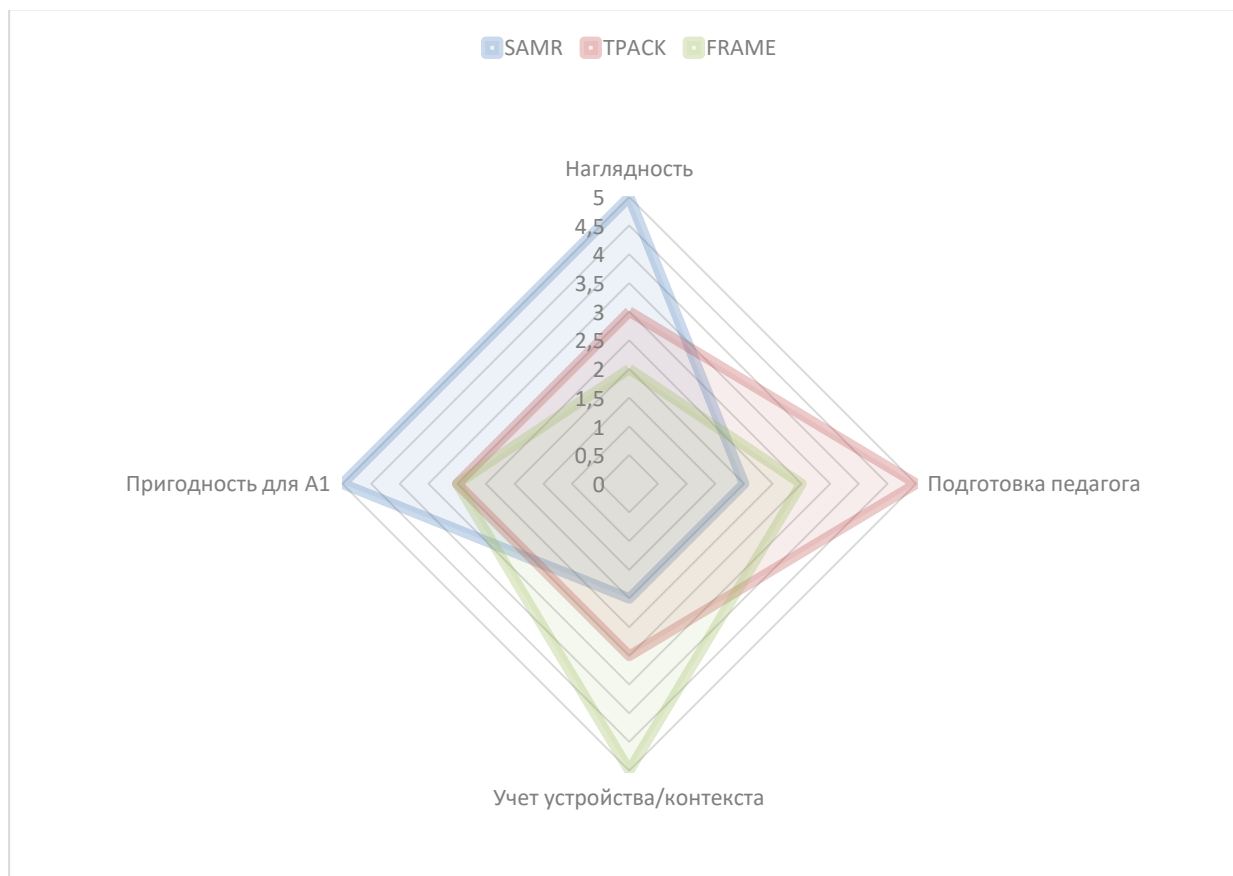


Рис. 1. Результаты сравнения моделей педагогами

Сравнительный анализ позволил сделать вывод о том, что в условиях преподавания РКИ на начальном уровне и при ограниченном времени эксперимента целесообразнее использовать модель SAMR как основную из-за ее практической направленности, позволяющей в короткие сроки внедрить AR-технологии без избыточной когнитивной нагрузки на студентов. Однако она, как и все рассматриваемые модели, нуждается в адаптации под задачи РКИ. В свою очередь модели TPACK и FRAME могут быть использованы с методической целью в качестве теоретической базы.

В качестве иллюстрации рассмотрим примеры заданий из урока по теме «Русская кухня», затрагивающие 1 и 2 ступени модели SAMR, так как студенты из Эфиопии испытывают затруднения в восприятии типично русских культурных понятий и ситуаций и владеют русским языком на достаточно низком уровне A1 (Рисунок 2).

Согласно разработанному в УЛГПУ учебно-методическому пособию для изучающих русский язык как иностранный «Приглашение в Россию», в рамках урока по данной теме студентам предлагается изучить 3-D модель самовара и других объектов по теме, прослушать аудиотекст про мороженое (1 ступень модели – замещение); получить интерактивные подсказки к модели самовара, прослушать мини-диалоги на употребление тематической лексики в авторском приложении, работающем на ОС Android (2 ступень модели – расширение (дополнение)).

Задание 5 (визуализация – самовар)

Место
для QR

Рассмотрите объект.
Ответьте на вопросы.

1. Как вы думаете, для чего нужна эта вещь?
2. Вы знаете, как называется эта вещь?
3. Прочитайте текст из Задания 6, чтобы узнать об этой вещи.

Задание 9 (аудиофайл – Русское мороженое)

Место
для QR

Послушайте текст, решите тест и ответьте на вопросы.

Рис. 2. Примеры заданий 1 ступени по модели SAMR

Такой подход, выбор модели SAMR, на практике подтвердил, что даже при ограниченном времени на занятия можно так методически строить уроки, что цифровые технологии выступают не как цель, а как средство, инструмент, способствующий активному формированию социокультурной компетенции через освоение лексики и погружение в культуру изучаемого языка. Дальнейшие исследования могут включать в себя попытки интегрировать все три модели с целью расширения методической базы, адаптацию под конкретные запросы РКИ, что позволит создать более персонализированные образовательные траектории и сценарии для изучающих русский язык как иностранный.

ЛИТЕРАТУРА

1. Fiorella, Logan & Mayer, Richard. (2021). Principles for Reducing Extraneous Processing in Multimedia Learning. DOI: <https://doi.org/10.1017/9781108894333.019>.
2. Features of using methods and means of the augmented reality technology when teaching a foreign language / N. A. Sergeeva, A. N. Zakharova, O. S. Rublyova, S. I. Tyutyunnik // Perspectives of Science and Education. – 2021. – No. 2(50). – P. 472–486. DOI: <https://doi.org/10.32744/pse.2021.2.33>.
3. Маркина П. В. Технологии AR и VR в обучении русскому языку как иностранному / П. В. Маркина, В. В. Маркин // Глобальный научный потенциал. – 2025. – № 2(167). – С. 134–137.
4. Puentedura R. R. (2013). SAMR: A model for technology integration. URL: http://www.hippasus.com/rrpweblog/archives/2013/03/05/SAMR_A_Model_for_Technology_Integration.pdf (дата обращения: 11.07.2025).
5. Koole M., McQuilkin J. L., & Ally M. Mobile Learning in Distance Education: Utility or Futility? // Journal of Distance Education / Revue de l'Éducation à Distance. – 2010. – Vol. 24, № 2. – P. 59–82. URL: <https://files.eric.ed.gov/fulltext/EJ892387.pdf> (дата обращения: 11.07.2025)
6. Hamilton E. R., Rosenberg J. M., Akcaoglu M. The Substitution Augmentation Modification Redefinition (SAMR) Model: A Critical Review and Suggestions for its Use // TechTrends. – 2016. – Vol. 60(5). – P. 433–441.
7. Shen Zhihua. (2021). An empirical study on adult students' English mobile learning based on frame model. E3S Web of Conferences. DOI: <https://doi.org/10.1051/e3sconf/202127502034>.

УДК 378.147:811

Зуевская Елена Викторовна

кандидат филологических наук,
доцент, заведующий кафедрой
теории и практики немецкого
языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Elena Zuevskaya

PhD in Philology, Associate Professor,
Head of the Department of Theory
and Practice of the German Language
Belarusian State University
of Foreign Languages
Minsk, Belarus
sujewska@yandex.by

Шинкаренко Дина Александровна

старший преподаватель кафедры
теории и практики немецкого
языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Dina Shinkarenko

Senior Lecturer of the Department
of Theory and Practice of the German
Language
Belarusian State University
of Foreign Languages
Minsk, Belarus
dinnaria@yandex.ru

**ОБУЧЕНИЕ ПИСЬМЕННОЙ РЕЧИ НА ПРОДВИНУТОМ ЭТАПЕ
С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА (немецкий язык)**

Современные тенденции в языковом образовании связаны с активным внедрением цифровых технологий, в частности искусственного интеллекта (ИИ), что открывает новые возможности для повышения эффективности обучения иностранным языкам. В статье рассматриваются возможности обучения письменной речи на продвинутом этапе с использованием технологий искусственного интеллекта. Описываются возможности ИИ-инструментов для формирования и совершенствования технических навыков письма, формирования языковой и текстовой компетенций, развития коммуникативно-прагматических умений, умений планирования, редактирования и рефлексии, а также обучение информационно-справочному поиску. На примере использования Peer AI Tutor в обучении студентов Минского государственного лингвистического университета демонстрируется эффективность ИИ в обучении написанию разножанровых текстов на продвинутом уровне владения немецким языком. Показано, что сочетание традиционных методов с цифровыми инструментами способствует повышению мотивации, снижению страха ошибок и достижению устойчивых образовательных результатов.

Ключевые слова: искусственный интеллект; письменная речь; языковое образование; цифровые технологии; текстовая компетенция.

**TEACHING ADVANCED WRITING SKILLS USING ARTIFICIAL
INTELLIGENCE TECHNOLOGIES (German language)**

Current trends in language education are connected with the active implementation of digital technologies, particularly of the artificial intelligence (AI), and this opens up new opportunities for improving the effectiveness of foreign language learning. The article discusses

the possibilities of teaching advanced writing skills using artificial intelligence. It describes the potential of AI tools for developing and improving technical writing skills, forming language and text competences, developing communicative and pragmatic skills, planning, editing and reflection skills, as well as reference search. The example of using Peer AI Tutor in teaching students at the Minsk State Linguistic University demonstrates the effectiveness of AI in teaching advanced-level German writing in various genres. It shows that combining traditional methods with digital tools helps to increase motivation, reduce fear of mistakes, and achieve sustainable educational results.

Key words: artificial intelligence; written language; language education; digital technologies; text competence.

Современные тенденции развития языкового образования тесно связаны с активным поиском инновационных образовательных технологий, направленных на повышение эффективности образовательного процесса по иностранному языку и достижению устойчивых образовательных результатов, а также со стремительным развитием цифровых технологий, в частности технологий искусственного интеллекта (ИИ). Популярность ИИ обусловлена его способностью обрабатывать огромные объемы данных быстрее и точнее, чем человек, возможностью использовать его для автоматического оценивания, персонализации и индивидуализации обучения, предиктивной аналитики.

В обучении иностранным языкам особую роль играет обучение письменной речи, которое предполагает формирование у обучающихся целого комплекса компетенций, необходимых для эффективной коммуникации в письменной форме [1, с. 308]. При этом обучение письменной речи включает в себя несколько ключевых аспектов:

1. Формирование и совершенствование технических навыков письма (начиная от правильного начертания букв и соблюдения правил орфографии и пунктуации до овладения базовыми нормами оформления и структурирования текста (абзацы, структура предложения, разделение текста на логические части и т.д.)).

2. Формирование языковой компетенции, которая подразумевает овладение лексико-грамматическими нормами языка, а также формирование навыков использования стилистически уместной лексики.

3. Формирование текстовой компетенции, т.е. развитие умения создавать связные, логически организованные тексты разных жанров (эссе, письмо, отчет, статья и др.), а также формирование навыка использования соответствующих средств связности.

4. Развитие коммуникативно-прагматических умений, способности учитывать адресата и цель письменного сообщения (например, убеждение, информирование, выражение эмоций) с учетом жанровых особенностей текстов (деловая переписка, академическое письмо, творческие тексты).

5. Развитие умений планирования, редактирования и рефлексии, а также обучение информационно-справочному поиску (использованию словарей, шаблонов, информационно-справочных систем, корпусов текстов, алгоритмов написания).

Далее будет предпринята попытка рассмотреть возможности различных технологических решений на базе ИИ для каждого из упомянутых аспектов.

Формирование технических навыков письма представляет собой фундаментальный этап, без которого невозможно дальнейшее обучение письменной речи. Он требует систематической тренировки, коррекции ошибок и интеграции с лексикой и грамматикой. К классическим методам и приемам формирования технических навыков относятся имитационные (списывание, письмо по образцу) и трансформационные упражнения (вставка пропущенных букв, исправление ошибок), написание диктантов, изложений, составление предложений по моделям. Современные технологии на основе ИИ могут оптимизировать данный процесс обучения, в частности компьютерное зрение (например, Google Handwriting Input, MyScript) преобразует рукописный текст в печатный, исправляя неточности, голосовые ассистенты (Google Assistant, Siri) и аналогичные приложения транскрибируют устную речь в текст, позволяя сравнивать написание с произношением. ИИ позволяет адаптировать диктанты, подбирая скорость и сложность диктовки под уровень обучающегося. Таким образом, технологии ИИ дополняют традиционные методы и приемы, не заменяя их и демонстрируя наибольшую эффективность при «смешанном формате обучения» [2, с. 1199].

Для формирования языковой компетенции можно использовать различные ресурсы для создания упражнений, направленных на тренировку и закрепление конкретного лексического и грамматического материала (MagicSchool, Twee, ChatGPT и др.), функциональные возможности которых позволяют создавать лексические и грамматические упражнения на заполнение пропусков, соотнесение слов с дефинициями, раскрытие скобок, расположение слов в правильном порядке и т.д. Многочисленные ресурсы могут исправлять ошибки, обеспечивая мгновенную обратную связь (ChatCPT, DeepL Write, YandexGPT, GigaChat и др.), поясняя, в чем заключается ошибка, и давая рекомендации по ее исправлению. Такого рода инструменты существенно облегчают работу преподавателя иностранных языков, позволяя сэкономить время на разработку упражнений и обеспечить персонализированный подход, а самостоятельная работа обучающихся за счет использования технологий ИИ имеет большую эффективность, «поскольку дает возможность получать мгновенную обратную связь, помогает студентам увидеть свои слабые стороны и ликвидировать пробелы, обладая при этом психологическим комфортом» [3, с. 136].

В обучении иностранным языкам особую сложность представляет собой формирование текстовой компетенции [4; 5]. Поскольку на данном этапе обучающиеся должны создавать связные, логичные и стилистически адекватные тексты, требуется одновременно контролировать грамматическую, лексическую стороны, учитывать структуру текста и коммуникативную цель, что требует высокого уровня автоматизации языковых навыков [5]. Ограниченный словарный запас и слабое владение грамматикой затрудняют построение сложных предложений и текстов. Особые сложности представляет собой создание связных текстов, с учетом когезии и когерентности.

Помимо этого тяжело дается учет жанровых особенностей различных видов письменных текстов. Как правило, такой вид работы как написание разножанровых текстов, осуществляется на продвинутом уровне изучения языка (B2 и выше) и подразумевает на первом этапе знакомство с языковыми и структурными особенностями конкретного жанра, их анализ, выявление типичных черт и языковых структур и лишь за этим следует этап продуцирования текста.

Соответственно сложной является также самопроверка такого рода текстов, поскольку такие ресурсы как ChatCPT, YandexGPT, GigaChat, DeepL Write могут дать рекомендации касательно лексики, грамматики и пунктуации, внести стилистические правки и предложить варианты формулировок для повышения ясности и убедительности, однако не всегда способны учесть жанровые особенности текстов. В таких случаях целесообразным является привлечение к работе специализированных ресурсов, созданных для проверки разножанровых текстов и соответствия их требованиям. Такого рода ресурсом является в частности Peer AI Tutor, созданный на базе Мюнхенского технического университета. Данный ресурс представляет собой инновационную платформу, предназначенную для развития умений письменной речи на немецком языке. Она сочетает в себе функции автоматической проверки, интерактивного обучения и аналитики прогресса.

На первом этапе задается тема и вид сочинения (эссе, доклад, аннотация, рассказ, рецензия, анализ и т.п.), выбирается вид школы (начальная, средняя, гимназия), класс, регион Германии. После этого обучающийся может напечатать свой текст, добавить его как картинку либо pdf-файл, указать при необходимости более развернутое задание, которое было дано преподавателем. Ресурс проверяет загруженный текст и дает свои рекомендации по его улучшению, особо отмечая те положительные стороны, которые есть у написанного текста. Кроме оценки содержания и языковой правильности, дается оценка логики и аргументации, заключения, наличия примеров и указания на источники. Таким образом, указанный ресурс позволяет студентам получить мгновенную обратную связь, ускоряя процесс обучения.

Использование указанного ресурса продемонстрировало высокую эффективность в обучении студентов специальности «Современные иностранные языки» в Минском государственном лингвистическом университете, когда студенты в рамках дисциплины «Язык средств массовой информации» учатся писать тексты различных жанров. После написания пробного эссе и его публикации с помощью ресурса «Форум» с LMS Moodle обучающиеся проводят взаимооценивание работ друг друга, после чего им предлагается ознакомиться с данными оценками и провести оценивание с помощью Peer AI Tutor. Как показывает опыт, студенты учитывают при дальнейших написаниях эссе и иных текстов как рекомендации, данные одноклассниками, так и рекомендации Peer AI Tutor, что позволяет существенно улучшить письменные тексты. Особо следует отметить, что Peer AI Tutor не просто редактирует текст либо исправляет ошибки, как иные ресурсы, упомянутые выше, а дает персональные рекомендации по совершенствованию

ванию текста, способствуя таким образом развитию метакогнитивных способностей. Благодаря быстрой обратной связи и возможности видеть прогресс, ИИ помогает преодолеть страх совершить ошибку и позволяет повысить уверенность в себе при письменном общении на иностранном языке, чтобы добиваться более значимых успехов.

Помимо обучения написанию текстов немаловажной задачей преподавателя иностранного языка на продвинутом этапе является обучение студентов использованию ресурсов для информационно-справочного поиска. Для студентов, изучающих немецкий язык, к минимально необходимому перечню таких немецкоязычных ресурсов относятся:

- онлайн-словарь [duden.de](https://www.duden.de);
- портал Совета по немецкому правописанию [rechtschreibrat.com](https://www.rechtschreibrat.com);
- Цифровой словарь немецкого языка [dwds.de](https://www.dwds.de), в частности, алгоритмы профиля слова DWDS-Wortprofil с сортировкой по частоте и с максимальной настройкой количества коллокаций;
- портал [wortschatz.uni-leipzig.de](https://www.wortschatz.uni-leipzig.de) с систематикой по идеографическому словарю немецкого языка Франца Дорнзайффа;
- одно- и многоязычные онлайн-словари, например, [langenscheidt.com](https://www.langenscheidt.com), краудсорсинговые [multitran.com](https://www.multitran.com) и dict.cc [6, с. 118].

Именно умение применять надежные ресурсы для информационно-справочного поиска в сочетании с умениями планирования, редактирования и рефлексии позволяет добиться максимальной эффективности в обучении письменной речи на продвинутом уровне. Такого рода использование технологий ИИ позволяет разнообразить иноязычную практику обучающихся и вывести процесс обучения на новый по степени решения когнитивных задач уровень [7, с. 10].

ЛИТЕРАТУРА

1. Титова, С. В. Обучение иноязычной письменной речи в цифровой среде вуза // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2023. Т. 28, № 2. С. 302–316. – DOI 10.20310/10.20310/1810-0201-2023-28-2-302-316.
2. Хаустов, О. Н. Обучение студентов написанию письменных работ на основе практики с инструментами генеративного искусственного интеллекта / О. Н. Хаустов, Т. Ю. Тормышова, Н. И. Суханова // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2024. Т. 29, № 5. С. 1194-1207. – DOI 10.20310/1810-0201-2024-29-5-1194-1207.
3. Зувская, Е. В. Искусственный интеллект в обучении иностранному языку: возможности и вызовы // Язык в эпоху цифровых трансформаций и развития искусственного интеллекта : сб. науч. ст. по итогам междунар. науч. конф., Минск, 23–24 окт. 2024 г. Минск : МГЛУ, 2024. С. 132–138.
4. Болотнова, Н. С. Текстовая компетенция и её роль в профильном обучении русскому языку // Русский язык в профильном обучении, Ярославль, 17–18 марта 2005 года. Том Часть 1. Москва ; Ярославль: Ремдер, 2005. С. 21–32.

5. Овчинникова, О. И. Формирование текстовой компетенции на родном (русском) и иностранном (английском) языках у учащихся старшей ступени обучения : автореферат дис. ... канд. пед. наук : 13.00.02. Орёл, 2011. 24 с.
6. Лойко, О. А. Обучение информационно-справочному поиску в преподавании письменного перевода // Материалы ежегодной научной конференции преподавателей и аспирантов университета, 16–18 апреля 2024 г. : в 2 ч. / Минский гос. лингвист. ун-т. Минск, 2024. Ч. 1. С. 117–119.
7. Сысоев, П. В. Технологии искусственного интеллекта в обучении иностранному языку // Иностр. языки в школе. 2024. № 3. С. 6–16.

УДК 372.881.111.1

Иваненко Леонид Алексеевич

учитель английского и немецкого
языка

ГБОУ гимназия № 631

Приморского р-на СПб

г. Санкт-Петербург, Россия

Leonid Ivanenko

Teacher of English and German

Grammar school 631 of Primorsky

district of the city of Saint-Petersburg

Saint-Petersburg, Russia

li1456182@gmail.com

Шавловская Дина Владимировна

кандидат филологических наук

учитель французского языка,

руководитель МО ИЯ

ГБОУ гимназия № 631

Приморского р-на СПб

г. Санкт-Петербург, Россия

Dina Shavlovskaya

PhD in Philology

Teacher of French,

head of FL department

Grammar school 631 of Primorsky

district of the city of Saint-Petersburg

Saint-Petersburg, Russia

dina.savlovska@gmail.com

Сыч Софья Алексеевна

учитель английского языка

ГБОУ гимназия № 631

Приморского р-на СПб

г. Санкт-Петербург, Россия

Sofia Sych

Teacher of English

Grammar school 631 of Primorsky

district of the city of Saint-Petersburg

Saint-Petersburg, Russia

bez.shutok@yandex.ru

ИСПОЛЬЗОВАНИЕ ИИ В ОБУЧЕНИИ АНГЛИЙСКОМУ ЯЗЫКУ В СРЕДНЕЙ ШКОЛЕ: АНАЛИЗ ДИАЛОГОВОГО ВЗАИМОДЕЙСТВИЯ

Современные цифровые технологии, включая искусственный интеллект (ИИ), активно трансформируют образовательные процессы, предлагая новые инструменты для развития языковых навыков. В данной статье исследуется потенциал ИИ-чат-ботов в обучении английскому языку на примере эксперимента с учащимися 6-х классов. Цель работы – оценить эффективность диалогового взаимодействия школьников с ИИ-со-беседником (DeerSeek Chat) в контексте развития коммуникативных компетенций. Методология включает дискурсивный анализ диалога между ученицей (12 лет, уровень CEFR A2) и чат-ботом на тему организации досуга. Анализ показывает, что ИИ успешно адаптируется к возрастным особенностям пользователя: использует упрощённый синтаксис, эмоциональные маркеры (смайлы) и тематически релевантные предложения. Однако выявлены ограничения, такие как неспособность бота распознавать скрытые смыслы, корректировать языковые ошибки и поддерживать сложные сценарии диалога. Таким образом, авторы приходят к следующим выводам: чат-бот эффективен для тренировки базовых коммуникативных навыков и снижения языкового барьера; управление диалогом остаётся за учеником, что способствует развитию инициативности; необходима доработка ИИ-алгоритмов для образовательных целей, включая функции коррекции ошибок и гибкого ветвления диалога. Практическая значимость исследования

заключается в рекомендациях по интеграции ИИ-инструментов в учебный процесс, учитывая их текущие ограничения. Перспективы включают изучение долгосрочного эффекта таких взаимодействий и модификацию чат-ботов для педагогических задач.

Ключевые слова: цифровые образовательные технологии; ИИ в образовании; английский язык; коммуникативные навыки; дискурсивный анализ.

THE USE OF AI IN TEACHING ENGLISH IN MIDDLE SCHOOL: THE ANALYSIS OF DIALOGUE INTERACTION

Modern digital technologies, including artificial intelligence (AI), actively transform educational processes, offering new tools for developing language skills. This article explores the potential of AI chatbots in teaching ESL using the example of an experiment with 6th grade students. The aim of the work is to evaluate the effectiveness of students' interaction with an AI interlocutor (DeepSeek Chat) in the context of the development of communicative skills. The methodology includes a discourse analysis of a dialogue between a student (12 years old, CEFR A2 level) and a chatbot on the topic of leisure activities. The analysis shows that AI successfully adapts to the age of the user: it uses simplified syntax, emotional markers (emoji) and thematically relevant suggestions. However, limitations have been identified, such as the bot's inability to recognize hidden meanings, correct language errors, and support complex dialogue scenarios. Thus, the authors come to the following conclusions: a chatbot is effective for training basic communication skills and reducing the language barrier; dialogue management remains with the student, which contributes to the development of initiative; AI algorithms need to be improved for educational purposes, including error correction and flexible dialogue branching functions. The practical significance of the study lies in the recommendations for integrating AI tools into the educational process, taking into account their current limitations. The prospects include studying the long-term effect of such interactions and modifying chatbots for pedagogical tasks.

Key words: digital educational technologies; AI in education; English as a second language; communication skills; discourse analysis.

Технологии значительно повлияли на все области нашей жизни, не могло остаться в стороне и образование. Интегрирование технологий в образовательный процесс позволили сделать передачу знаний быстрее, более удобной и эффективной, к тому же, это отвечает интересам обучающихся, ведь современные школьники стремятся применять технологии везде, где это возможно, в том числе в процессе обучения. Огромную роль в процессе обучения играет интернет. Он содержит значительное количество образовательных ресурсов, служит важнейшим источником информации.

Цифровые инструменты являются двигателем прогресса в современной образовательной мысли. Инновационная деятельность в образовании так или иначе связана с новыми технологиями. Адаптивность и свобода, которую предлагают цифровые методы, делает обучение более притягательным для младшего поколения [1, с. 76–101].

Тем не менее введение новых технологий в образование не всегда проходит беспрепятственно с самого начала. Не изжил себя взгляд на техно-

логии как на средство скорее развлечения, нежели обучения, поэтому каждое нововведение проходит определённый период адаптации, прежде чем перейти в широкое использование.

Относительной новинкой цифрового мира является искусственный интеллект (далее – ИИ). ИИ представляет собой технологию, которая позволяет компьютерной программе имитировать человеческое обучение, понимание, самостоятельное принятие решений и творчество. Он может понимать человеческий язык и даже отвечать [2]. Последнее особенно важно на уроках иностранного языка.

Исследователи не обошли стороной этот вопрос, но можно сказать, что это достаточно молодая область исследования. Целью научных изысканий становились всевозможные ресурсы на основе ИИ, а также их применение в различных областях преподавания. Первые исследования были сделаны в области автоматизированной обработки письменных работ. Шермис и Бёрстайн выяснили, что ИИ может дать обратную связь о грамматике, стиле и логичности текста [3]. Уилсон и Чжик определили, что ИИ, несомненно, снижает нагрузку на учителя, тем не менее некоторые аспекты оценивания для него могут быть пока не под силу, например, творческая составляющая или авторский стиль [4]. В работах Дива и др. утверждается, что обучающиеся могут учиться быстрее, опираясь на ИИ, но могут полагаться на него слишком много. Также, написание промптов и совместное с ИИ творчество позволяет повысить креативность обучающихся [5]. Исследованием возможностей ИИ для чтения занимались Хуань и Тан. ИИ способен адаптировать тексты для чтения, чат-боты можно использовать для создания вопросов к текстам, что касается текстов, написанных самим ИИ, здесь присутствует высокий риск плагиата и отсутствия оригинальности [6].

Обширные исследования были сделаны в области использования ИИ при обучении говорению. Так, было установлено, что ИИ позволяет значительно улучшить произношение [7], в то же время, отмечали, что инструменты на базе ИИ более эффективны при работе с сегментными единицами произношения [8]. Использование чат-ботов в практике говорения позволило обучающимся улучшить собственные навыки ведения беседы и повысить уверенность, но при этом их смущала неестественная манера общения виртуального собеседника. Другие исследователи также отмечали, что ИИ не может с полной достоверностью передать манеру общения настоящего человека. Согласно исследованиям Дайзона, наибольшую пользу из простого диалога с ИИ смогли получить обучающиеся с более низким уровнем языка, в то время как продвинутые студенты нуждались в большей сложности [9].

Развитие технологий не стоит на месте, и новые инструменты продолжают появляться, а те, которые уже есть, постоянно обучаются. Отсутствие качественной аутентичной коммуникативной практики для обучающихся

средних общеобразовательных школ ставит ряд вопросов: насколько продуктивной может быть организованная беседа школьников с ИИ, какова структура и каковы особенности диалога с виртуальным собеседником? Ответы на эти вопросы составили цель настоящей статьи.

Методология

Для того, чтобы найти ответы на эти вопросы, обозначенные выше, мы провели эксперимент на базе гимназии. Эксперимент проходил с участием обучающихся шестых классов гимназии № 631 Приморского района Санкт-Петербурга. В эксперименте приняли участие 25 человек. Эксперимент проходил в компьютерном классе, под наблюдением учителя. Задание: разговор с ИИ. Тема: занятия в свободное время. Задача, которая стояла перед обучающимися, была сформулирована следующим образом: им необходимо было ввести промпт, заранее разработанный и написанный на доске, и вступить в диалог с ИИ. Диалог должен был включать по меньшей мере 10 реплик со стороны обучающегося. В качестве собеседника был выбран чат-бот от компании DeepSeek, главным образом, ввиду своей доступности и возможности сохранения чатов в профиле пользователя. Продолжительность эксперимента: 45 минут.

Подготовительным этапом для данного эксперимента выступило аналогичное задание в классе, в ходе которого в роли консультантов по свободному времяпрепровождению выступали педагоги английского языка. Во время беседы, обучающиеся должны были ответить на предложения о том, чем заняться в свободное время, и согласиться или не согласиться. Тема была выбрана с учётом возрастных особенностей и уровня английского языка обучающихся с тем расчётом, чтобы они смогли вести диалог на приемлемом для понимания исследователя уровне.

Поскольку киберкоммуникация существует в текстовой форме и обладает акциональной природой, её анализ требует методов, направленных одновременно на изучение текстовых феноменов и социальных действий. В свете этих факторов было принято решение остановиться на методах дискурсивного анализа. В данной статье предпринята попытка подробно проанализировать один диалог.

Анализ интеракций

Участники: Таисия (T), чат-бот (D).

T: i need advice on what to do at the weekend. i'm 12 years old and go to school in Saint-Petersburg, Russia. Tell me what i could do at the weekend. Keep your answer short and simpl. Give me one option at a time.

D: Option 1: Visit the Hermitage Museum-it's huge and full of cool art and history. You can explore for hours! Want another idea? Just ask!

T: Hello my name is Taisia

D: Hi, Taisia!

Option 2: Go to Peterhof Palace (if it's open)-see the amazing fountains and gardens. It's like a royal summer playground! Want more? Let me know!

T: yes, good idea

D: Great! Option 3: Explore St. Isaac's Cathedral-climb to the colonnade at the top for an awesome view of the whole city! Next idea? Just say!

T: yes

D: Option 4: Visit Lenpark or Tauride Garden-rent a bike, have a picnic, or just relax in nature. More? Let me know!

T: wow, i like rent a bike

D: That's awesome! Option 5: Bike along the Neva River embankments (like the Palace Embankment) – you'll see cool bridges, ships, and palaces while riding! Another idea? Just ask!

T: oh, i want drink coffee. go to the cafe?

D: Option 6: Try "Double B" or "Ziferburg" cafés-they have yummy drinks and cozy vibes. Perfect for a relaxing break! More ideas? Just say!

T: ok

D: Option 7: Visit "Bookvoed" (Буквоед) on Nevsky Prospekt-it's a cool bookstore with a café inside. Read, sip hot chocolate, and browse books! Next? Let me know!

T: well, i buy a manga

D: Option 8: Check out "Comic Street" or "Labyrinth Bookstore"-they have great manga sections! Grab a new volume and enjoy. More? Just ask!

T: no, i don't enjoy

D: No problem! Option 9: Try arcades or VR gaming (like "Virtual Reality Club")-play fun games and escape into a cool digital world! Want something else?

T: at home?

D: Option 10: Have a movie marathon-pick a fun series (like Studio Ghibli or Harry Potter), make popcorn, and cozy up! Or want a creative idea?

T: well

D: Option 11: Try baking cookies or cupcakes-easy recipes online, and you get a tasty reward! (Bonus: share with family/friends!) Or want a lazy-day idea?

T: yes. i love it

D: Option 12: Build a blanket fort in your room-add pillows, fairy lights, and watch movies or read manga inside!

Perfect lazy-day fun. More ideas? Just ask!

T: i interested it. what manga we will read?

D:Option 13: Try these manga for a cozy blanket fort session:

- "My Neighbor Totoro" (chill & magical)
- "Yotsuba&!" (funny slice-of-life)
- "Spy x Family" (action + humor)

Grab snacks and dive in!

Need more picks? Just ask!

T: okay, my neighbor totoro

D: Great choice! Option 14: Make it extra cozy-pair "My Neighbor Totoro" with:

- Studio Ghibli soundtrack in the background
- Totoro-shaped snacks (try cutting apples or pancakes like him!)

Total immersion! Want to add another Ghibli movie?

T: no. i don't want

D: Got it! Option 15: Switch gears-try drawing your own manga character! Grab pencils and invent a cool hero or creature. Or want a zero-effort idea?

T: yeah. the best idea

Анализ коммуникативных стратегий в диалоге между чат-ботом и учащейся выявляет несколько значимых аспектов взаимодействия. Прежде всего, наблюдается успешная адаптация искусственного интеллекта к возрастным особенностям собеседника, что соответствует современным принципам age-appropriate design в образовательных технологиях [10]. Использование смайликов со стороны бота выполняет важную прагматическую функцию, способствуя созданию дружелюбного и доверительного коммуникативного пространства. Также, использование эмодзи соответствует принципам мультимодальной коммуникации, создавая эффект социального присутствия.

Помимо этого, возрастная адаптация бота проявляется в:

- упрощённом синтаксисе;
- тематическом выборе;
- эмоциональном оформлении реплик.

Обучающаяся, в отличие от типичных паттернов peer-communication, сознательно избегает применения аналогичных визуальных средств экспрессии.

Эмоциональная вовлеченность респондента проявляется через систему междометных реакций (*wow* 'вау', *oh* 'оу') и кратких эмотивных высказываний, что свидетельствует о формировании определенного уровня психологической связи с цифровым собеседником. Примечательно, что управление диалогической структурой полностью осуществляется студенткой, которая выступает инициатором всех тематических переходов. Присутствует чёткое распределение ролей в диалоге: обучающая как инициатор, чат-бот – респондент. При этом чат-бот применяет двухуровневую стратегию реагирования на отказ: при нейтральной реакции (*ok* 'окей', *well...* 'хорошо') происходит углубление аргументации, при этом предложение не становится абсолютно новым, но немного видоизменяется. При негативной реакции на предложение бота (*I don't want* 'не хочу') происходит полный тематический сдвиг.

Лаконичность реплик обучающейся может интерпретироваться в двух аспектах: как проявление индивидуальных коммуникативных особенностей, так и показатель текущего уровня владения языковыми компетенциями (CEFR A2). Также, преобладание коротких ответов может свидетельствовать о:

- трудностях в преодолении языкового барьера;
- настороженности по отношению к ИИ-собеседнику;

– особенностях учебной коммуникации.

Важным наблюдением является нейтральная позиция чат-бота по отношению к языковым ошибкам, что объясняется отсутствием явно заданной функции коррекции в его алгоритмах. Тем не менее техническая архитектура подобных систем потенциально допускает их модификацию для целей языкового супервизорства через промпт [11].

Заключение

Проведённый анализ говорит о том, что ИИ имеет определённый потенциал как инструмент для развития базовых коммуникативных навыков, снижения языкового барьера и безопасной речевой практики на иностранном языке. В то же время, представленные стратегии реагирования на ответы обучающихся демонстрируют неспособность чат-бота распознавать скрытые смыслы ведёт к ограничению сценарного ветвления диалога и приводит к однообразию реакций, что может вызвать скуку и обучающихся, и должно учитываться при моделировании учебных заданий.

Перспектива подобных исследований состоит в возможности изучать эффект от подобных речевых взаимодействий в долгосрочной перспективе. Также, есть возможность изучить педагогический потенциал путём активизации корректирующих алгоритмов у чат-бота. Выявленная интертекстуальность может быть источником дополнительной мотивации обучающихся к изучению английского языка, что также требует дальнейшего изучения.

ЛИТЕРАТУРА

1. Розенче Дж., Пиа Р., Ходли К. Будущее цифрового обучения в школе / Дж. Розенче, Р. Пиа, К. Ходли. – Москва: Просвещение, 2000. – 76–101 с.
2. IBM. What is Artificial Intelligence? [Электронный ресурс] // IBM Think. – 2023. – URL: <https://www.ibm.com/think/topics/artificial-intelligence> (дата обращения: 10.07.2024)
3. Shermis M. D., Burstein J. C. Handbook of Automated Essay Evaluation: Current Applications and New Directions. – New York: Routledge, 2013. – 456 p.
4. Wilson J., Czik A. Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality // Computers & Education. – 2016. – Vol. 100. – P. 112–126.
5. Deeva G., Bogdanova D., Serral E., Snoeck M., De Weerd J. A review of automated feedback systems for learners: Classification framework, challenges, and opportunities // Computers & Education. – 2021. – Vol. 162. – P. 94–104.
6. Huang X., Tan S. The role of AI chatbots in fostering L2 English learners' reading comprehension and engagement // Artificial Intelligence in Education: 24th International Conference, AIED 2023, Proceedings. – 2023. – P. 215–227.
7. Ahn J., Lee K., Kim B. The impact of AI pronunciation feedback on L2 learners' intelligibility // ReCALL. – 2021. – Vol. 33, No 3. – P. 327–343.

8. Chen L., Wang N., Zhao H. Automated feedback in second language speaking practice: A meta-analysis // *Language Learning & Technology*. – 2022. – Vol. 26, No 1. – P. 1–24.
9. Dizon G. Evaluating intelligent personal assistants for L2 listening and speaking practice // *Journal of Asia TEFL*. – 2020. – Vol. 17, No 3. – P. 1028–1044.
10. Livingstone, S., Stoilova, M., Nandagiri, R. (2021). "Children's data and privacy online: Growing up in a digital age". *New Media & Society*, 23(8), 1169–1186. DOI: <https://doi.org/10.1177/14614448211020544>
11. Warschauer M. Digital inequality and language learning: A critical review // *System*. – 2020. – Vol. 92. – P. 102271. DOI: 10.1016/j.system.2020.102271.

УДК 811.1/.8(072):004

Ковальчук Алина Михайловна

старший преподаватель кафедры
английского языка гуманитарных
факультетов
Белорусский государственный
университет
г. Минск, Беларусь

Alina Kovalchuk

Senior lecturer at the Department
of English for Humanities
Belarusian State University
Minsk, Belarus
alina.kovalchuk.2011@mail.ru

Кудырко Ирина Владимировна

старший преподаватель кафедры
английского языка гуманитарных
факультетов
Белорусский государственный
университет
г. Минск, Беларусь

Irina Kudyrko

Senior lecturer at the Department
of English for Humanities
Belarusian State University
Minsk, Belarus
KudyrkaIV@bsu.by

Трубочкина Инна Анатольевна

старший преподаватель кафедры
английского языка гуманитарных
факультетов
Белорусский государственный
университет
г. Минск, Беларусь

Inna Trubochkina

Senior lecturer at the Department
of English for Humanities
Belarusian State University
Minsk, Belarus
innatrubochkina@mail.ru

РОЛЬ ЭЛЕКТРОННЫХ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ В ОРГАНИЗАЦИИ ВОСПИТАТЕЛЬНОЙ РАБОТЫ ПОСРЕДСТВОМ ИНОЯЗЫЧНОГО ОБРАЗОВАНИЯ

В данной статье рассматриваются возможности использования электронных образовательных ресурсов при организации воспитательной работы в процессе преподавания профессионально ориентированного английского языка в неязыковом вузе. Подчеркиваются преимущества последних для создания новой образовательной среды при формировании практико-ориентированных компетенций, а также акцентируется внимание на возможностях образовательного портала БГУ при проведении воспитательной работы с учащимися. В публикации приводятся практические примеры реализации технологии смешанного обучения в ходе внедрения иноязычных образовательных и воспитательных проектов, направленных на формирование цельной языковой личности студента. Рассматриваются особенности проведения традиционных Недель английского языка посредством внедрения цифровых ресурсов. Авторы наглядно иллюстрируют потенциал цифровизации в организации воспитательной работы в процессе иноязычного образования. Также в статье отражены методические приемы, используемые при проведении разнообразных воспитательных мероприятий на иностранном языке, базирующиеся на применении цифровых ресурсов.

Ключевые слова: воспитательная работа; электронные образовательные ресурсы; критическое и рефлексивное мышление; диджитализация образовательной среды; профессиональные компетенции; междисциплинарная кооперация.

DIGITAL EDUCATIONAL RESOURCES FOR ORGANIZING EXTRA-CURRICULAR ACTIVITIES IN TEACHING FOREIGN LANGUAGES

This article discusses the advantages of using digital educational resources for organizing extracurricular activities in the process of teaching professionally oriented English in a non-linguistic university. The advantages of the latter for creating a new educational environment in the formation of practice-oriented competencies are emphasized, and attention is also focused on the possibilities of the BSU educational portal while conducting extracurricular activities with students. The publication provides practical examples of the implementation of blended learning technology during the implementation of foreign language educational and upbringing projects aimed at forming a holistic linguistic personality of the student. The features of holding traditional English Language Weeks through the introduction of digital resources are considered. The authors clearly illustrate the potential of digitalization in organizing educational work in the process of foreign language education. The article also reflects the methodological techniques used in conducting various educational activities in a foreign language, based on the use of digital resources.

Key words: extracurricular educational activities; electronic educational resources; critical and reflective thinking; digitalization of the educational environment; professional competencies; interdisciplinary cooperation.

Воспитательная работа в современной системе высшего образования является важнейшим фактором формирования конкурентоспособной самодостаточной личности, обладающей профессиональными знаниями и умениями, высоким культурным уровнем, чувством ответственности, высокой степенью нравственности, способную к самосовершенствованию. Для достижения всех поставленных целей важно создать единую систему воспитательной работы и выработать единые подходы к ее проведению, а также возможности ее интеграции в образовательный процесс. Существование ряда проблем, возникающих с привлечением студентов к участию в общественной жизни вуза, факультета, своей специальности, ставит перед преподавателями необходимость переосмысления подходов и внедрение новых форм воспитательной работы. Кроме того, в современных условиях диджитализации образовательной среды, воспитательная работа также претерпевает ряд изменений. Появилась возможность активно использовать электронные образовательные ресурсы (ЭОР) для качественного улучшения внеучебной деятельности студентов. ЭОР способны обеспечить интерактивность образовательного процесса, повышают интерес студентов к дисциплине благодаря наглядности, доступности, разнообразию и привлекательности визуального контента, а также возможности выбора актуальной информации по теме исследования. Свою задачу преподаватели кафедры английского языка гуманитарных факультетов БГУ видят в том, чтобы в максимальном объеме

обеспечить образовательные потребности учащихся на занятиях в аудитории и во внеучебной деятельности, превратить их способности в достижения и помочь им реализоваться в дальнейшей профессиональной жизни, активно участвовать в межкультурной коммуникации, реализовать свои личностные планы. Решение данных задач достаточно успешно осуществляется, на наш взгляд, в рамках ряда мероприятий, проводимых на экономическом факультете в течение всего курса обучения английскому языку, среди которых одним из самых масштабных является организация и проведение мероприятий ежегодной Недели английского языка.

Условия вынужденного перехода в 2020 г. в режим дистанционного обучения мотивировали преподавателей кафедры организовать проведение традиционной Недели английского языка на образовательном портале БГУ в онлайн-формате. Сочетание ресурсных возможностей портала, активное привлечение ЭОР и традиционных подходов в обучении и воспитании кардинально изменило образовательный процесс, а также характер сотрудничества преподавателя и студента. Создание портала БГУ позволило размещать проекты, презентации, доклады и другие творческие работы учащихся, а также информационно-методическое сопровождение для их выполнения в свободном доступе для ознакомления, а также по итогам Недели организовать онлайн обсуждение на форуме и проведение онлайн голосования студентов за лучшие проекты, организовать рефлексивные форумы. Возможность участвовать в мероприятиях Недели в онлайн режиме значительно расширяет круг ее участников и вариативность предлагаемых форматов участия. Таким образом, происходит расширение образовательных возможностей учащихся за счёт увеличения доступности и гибкости образования, учитываются их индивидуальные образовательные потребности. Во время проведения Недели студенты 1–2 курсов экономического факультета имеют возможность проявить свои самые разнообразные умения и реализовать самые нереальные проекты и замыслы. В каждом академическом году она проходит под разными девизами, которые определяют общую тематику Недели. Например, «*My Native Land, I Love You Sweet!*», «*Bright Year, Bright Me!*», «*Preserving the Past for the Future*» и т.д.

Программа Недели предлагает проведение целого ряда мероприятий как в аудитории, так и в онлайн-формате на общеуниверситетской образовательной платформе LMS Moodle. В рамках Недели английского языка проводится целый ряд мероприятий, направленных на расширение культурологического и профессионального кругозора студентов, развитие их научно-исследовательского и творческого потенциала, эстетического вкуса, укрепление междисциплинарных связей. Самым значимым событием, неизменно вызывающим огромный интерес у студентов, является конкурс на лучший стартап проект, который проходил под девизом «*Think Big, Dream Even Bigger!*». Ежегодно на суд зрителей и жюри студентами экономического

факультета представляется огромное количество бизнес-проектов. Работы учащихся строятся на глубоком знании структуры и содержания финансового проекта и предлагаемыми бизнес идеями. Кроме того, нестандартность подходов в реализации своих идей, их определенное новаторство, умение убеждать потенциальных инвесторов, блестящее владение языком и умение публичного выступления – вот далеко не все учебные и профессиональные умения и навыки, которые неизменно демонстрируют конкурсанты. Перед зрителями разворачивается реальная битва инновационных проектов, в которой победителей оказывается гораздо больше, чем побежденных, так как каждая презентация есть возможность попробовать свои силы в защите социально значимого продукта. Студенты неизменно демонстрируют огромный энтузиазм и желание убедить в полезности и актуальности своей идеи не только сверстников, но и более искушенную публику в лице преподавателей иностранного языка, а также преподавателей экономических дисциплин и студентов старших курсов – победителей и участников обще факультетской конференции лучших стартапов. Выбор данного формата далеко не случаен. Следует отметить некоторые возможности и преимущества, которые он предоставляет его участникам. Во-первых, это проект, который продвигает инновационную идею или содержит усовершенствованную модель старой, т.е. у его создателей есть свобода выбора не только идеи для нового бизнеса, но и способа ее публичной защиты. Во-вторых, для тех учащихся, которые умеют нестандартно мыслить, обладают аналитическими умениями, стремятся к самореализации и самосовершенствованию, данный формат внеучебной деятельности предоставляет такую возможность. В-третьих, студенты осуществляют познавательно-исследовательскую деятельность в рамках нескольких учебных предметов, что расширяет их возможности и убеждает в необходимости межпредметной кооперации, тесном общении с преподавателями дисциплин профессионального цикла. Кроме того, учащиеся реализуют свой творческий потенциал. Создается ситуация успеха и стимула для дальнейшего творчества. Работа над бизнес-проектами, докладами, выступлениями на конференциях или защита своей точки зрения за круглым столом носят развивающий характер. Усилиями преподавателя и наиболее мотивированных учащихся создается такая образовательная среда, в которой усиливается речемыслительная активность учащихся, имеет место проявление их критического и рефлексивного мышления. Такая возможность появляется у студентов на этапе публичной защиты своих бизнес-проектов. Создание такого проекта является открытым эвристическим заданием. Получаемый результат всегда уникален и отражает степень индивидуального творческого самовыражения каждого студента и группы в целом. Открытое задание, наполненное личностно-значимым для обучающегося смыслом, обладает высоким мотивационным потенциалом, способствует более глубокому

усвоению изучаемой дисциплины. Как и любое эвристическое задание, создание любого проекта должно начинаться с нестандартной идеи. Студенты самостоятельно формируют рабочие группы, распределяют обязанности, продумывают какие материалы будут использованы. Далее каждая группа осуществляет смысловое наполнение проекта, выбирает форму его реализации и его оформление. Роль преподавателя заключается в консультировании учащихся по соблюдению определенных критериев экономической составляющей проекта, выработанных коллегиально и согласованных с преподавателями дисциплин по специальности. Преподаватель определяет временные рамки проекта, рекомендует материалы и наиболее адекватные сайты для изучения. Кроме того, все вспомогательные материалы размещены на портале смешанного и дистанционного обучения (LMS Moodle) [1], который используется как для обеспечения учебного процесса, так и для организации воспитательной работы.

Следующим знаковым событием Недели для студентов 1 курса стал экономический форум «Экономика во имя всеобщего блага» («*Economics for the Common Good*»). Студенты выступили с докладами о проблемах современной экономики и путях их решения. После выступления участников, у зрителей была возможность обсудить проблемные вопросы за круглым столом, а также высказать своё мнение по поводу наиболее значимых глобальных проблем, таких как безработица, бедность, качество и уровень жизни в разных странах и другие не только в аудитории, но и онлайн. По окончании обсуждения всем участникам форума предоставляется возможность анонимно с помощью средств цифрового голосования выделить наиболее актуальные проблемы и вопросы для дальнейшего рассмотрения и обсуждения. Данный формат позволяет студентам развивать нестандартность и критичность мышления, стимулирует развитие таких основных качеств креативности, как беглость, гибкость оригинальность мысли, активное творческое саморазвитие, интеллектуальную самостоятельность, открытость к приобретению нового опыта. При поиске фактической информации, анализе и сличении полученных данных, освещенных в докладах выступающих, использовались цифровые ресурсы наряду с приложениями для подготовки цифровых презентаций (Canva, Gamma, Google Slides). Данные приложения имеют широкие возможности для создания графиков, анимации, звуковых эффектов, они помогают задействовать особенности искусственного интеллекта в разумной и сбалансированной пропорции.

Самые мотивированные и уверенные в себе студенты экономического факультета пробуют свои силы в конкурсе ораторского мастерства «*Public Speaking Contest*». Предварительно студенты готовят свои профили, в которых знакомят публику со своими интересами и объясняют выбор темы выступления. На подготовительном этапе все материалы конкурса, включая критерии оценивания, тематику выступлений, профили участников размещаются на образовательном портале кафедры, что позволяет сделать всю информацию доступной, привлечь внимание к конкурсу, сделать его зна-

чимым событием для других студентов. Заявленная тематика ранжируется по категориям: 1) самореализация и профессиональный рост; 2) критические вопросы высшего образования; 3) защита окружающей среды; 4) информационно-коммуникационные технологии; 5) будущая профессия. Организационный комитет разрабатывает критерии оценивания выступлений. Жюри оценивает выступления участников на предмет содержания и актуальности проблемы, новизны исследования, уровня риторических навыков, умения решать коммуникативную задачу и уровня владения иностранным языком. Данное мероприятие, на наш взгляд, значительно повышает интерес учащихся к исследовательской деятельности и к изучению иностранного языка в целом. И то, и другое требует определенных волевых усилий, навыков саморегуляции, что является проблемой для большинства учащихся. Привлекая учащихся к исследовательской деятельности, мы помогаем им преодолеть себя, формируем и развиваем способность ставить перед собой задачи, намечать пути их достижения, планировать, анализировать и рефлексировать свою деятельность. В индивидуальном плане, происходит повышение самооценки и более глубокое осознание возможности использования иностранного языка, как инструмента познания и средства общения.

Среди онлайн-форм Недели следует отметить виртуальный литературный салон «Слова без границ» («*Words without Borders*»), целью которого является развитие не только эстетического вкуса студентов, но и ряда умений ораторского мастерства, таких как убедительность, умение удерживать внимание аудитории, владеть своим голосом. Студенты размещают на платформе свои любимые поэтические произведения и пробуют свои силы в переводе этих произведений на английский язык и наоборот, создают постеры личного эмоционального восприятия произведения. Участники проявили интерес к созданию видео роликов собственной декламации любимых литературных произведений. Условием для чтецов являлось не просто внятное чтение, а интерпретация стихотворения, передача его смысла и настроения через голосовую модуляцию, выразительность и индивидуальную манеру декламации. Условием при оценивании и выборе лучших чтецов являются художественная выразительность и интонационное оформление поэтического произведения. Литературный онлайн-форум «Книги, открывшие эру нашей жизни» («*Books Which Opened the Era of our Life*»), становится логичным завершением события и замечательной возможностью для преподавателей провести рефлексию результатов мероприятия, а для ребят поделиться впечатлениями о прочитанных книгах и порекомендовать их другим, а также поделиться своими размышлениями и критическими замечаниями.

Один из дней Недели традиционно посвящается развитию общего кругозора, эрудиции, смекалки и гибкости ума. Рубрика «*Think Outside the Box*» неизменно пользуется популярностью у студентов. Десятки квизов, интерактивных игр и марафонов, создаваемых студентами и размещаемых на сервисе *drive.google.com* погружают ребят в увлекательный мир событий, фактов, уникальных открытий и просто забавных идей и изобретений. Но

кульминационным моментом для самых смекалистых и эрудированных студентов является командная интерактивная интеллектуальная игра «*NonTrivial Big Fun Game*». Каждое новое задание для команд отличается от предыдущего новизной, оригинальностью и свежестью идеи. Создатели игры наполняют ее содержанием и, в то же время, весельем и здоровой конкуренцией стать лучшими из лучших. Такого рода задания способствуют индивидуализации образовательного процесса. Каждый учащийся самостоятельно определяет свои задачи и цели, осуществляет поиск и отбор релевантного материала в интернете и его визуальное оформление с учетом своих образовательных потребностей, интересов и способностей. При этом студенты широко используют возможности различных образовательных сайтов, интерактивных платформ, онлайн-сервисов для создания увлекательного и познавательного контента.

Не менее интересным и мотивирующим мероприятием, по мнению его организаторов и участников, признан «Лекционный марафон», основной целью которого является вовлечение студентов в формат сотрудничества с преподавателями не только английского языка, но и профильных дисциплин. Командная работа по разработке и проведению в аудитории лекции по тематике курса с применением электронных образовательных ресурсов на английском языке превращается в творческую площадку по подбору материала для лекции сети Интернет и других источниках, его анализу с точки зрения релевантности, актуальности и новизны. На подготовительном этапе группа самостоятельно подбирает и обсуждает материал, структурирует его, логично выстраивает, выбирает адекватные формы представления с использованием цифровых инструментов, продумывает приемы интерактивной подачи для своей аудитории. Критические замечания преподавателя по профильной дисциплине заключаются в его экспертной оценке познавательной ценности материала и его объема для реализации поставленной цели. Его роль носит консультационный и корректирующий характер. Ход лекции, ключевые вопросы, терминология на английском языке, необходимые образительные содержательные и смысловые опоры записываются на электронный носитель и заранее размещаются на образовательной платформе с целью предварительного ознакомления слушателей, что снимает или минимизирует ряд трудностей, связанных с восприятием, пониманием и осмыслением материала. Кроме того, студенты, которые отсутствуют на лекции, имеют возможность виртуально принять участие в ней, так как в лекционных аудиториях предусмотрена такая возможность для студентов. Данный формат привлечения учащихся к внеучебной деятельности в новой для них роли имеет ряд преимуществ. Совершенствуется целый спектр профессиональных и коммуникативных компетенций.

В рамках Недели также успешно проходят мероприятия по повышению культуры поведения студентов, их способности общаться и проводить свой досуг. Они направлены на развитие эмоциональной сферы личности, необходимости следования этическим нормам. Студентами готовят и успешно

проводят праздник юмора «*Merry-making Party*». Огромную изобретательность, смекалку и творческий подход проявляют организаторы праздника. Участники делятся забавными традициями празднования Дня юмора в разных культурах, особенностями британского юмора, решают остроумные задачки, разыгрывают забавные скетчи. Интернет предоставляет богатейший материал для создания творческих проектов, сценического воплощения идей, создание интерактивных развлекательных викторин и конкурсов. Студенты фонтанируют своей изобретательностью, находчивостью, нестандартностью мышления, разнообразием своих интересов и талантов.

Еще одним знаковым мероприятием стало ежегодное участие студентов в международных олимпиадах по иностранному языку (Нижевартовск, Санкт-Петербург и др.) и предоставляет участникам уникальные возможности для развития своих талантов. Этапы проведения таких мероприятий сегодня организованы в основном онлайн.

Проведение языковых олимпиад в лингвистическом вузе и подготовка студентов для такого рода языковых соревнований требует качественно новой организации. Цифровые ресурсы и ЭОР кафедры являются своеобразной образовательной площадкой, призванной систематизировать данную работу, выстроить образовательный маршрут, разработать стратегию языковой подготовки и обучающие материалы. Такая форма построения учебного процесса позволяет участникам на совершенно новом уровне реализовать себя, развивая свои профессиональные лингвистические компетенции. Благодаря цифровому формату в виде отборочных асинхронных туров и собеседований финалистов в синхронном онлайн режиме, олимпиады способствуют налаживанию контактов среди студентов и преподавателей разных университетов на международной арене, популяризируя высшее учебное заведение-организатора олимпиады в студенческих и научных кругах.

Подводя итоги вышесказанному, следует отметить, что в ходе подготовки и реализации каждого мероприятия, студенты совершенствовали наиболее значимые профессиональные качества и компетенции, а именно, критическое и рефлексивное мышление, умение анализировать информацию и адаптироваться к меняющимся условиям, решать сложные проблемы, взвешивать альтернативные мнения и отстаивать свою точку зрения. Ярко проявились организаторские и коммуникативные умения многих учащихся, а также чувство ответственности и самостоятельности. Следует отметить, что использование цифровых технологий кардинально меняет весь процесс обучения. Происходит адаптация учебной деятельности к индивидуальным потребностям учащихся, повышается их вовлеченность и уровень мотивации к изучению иностранного языка, упрощаются организационные задачи при подготовке к занятиям и воспитательным мероприятиям. Цифровые инструменты обеспечивают обратную связь в реальном времени, что помогает быстрее и эффективнее провести мероприятие, выявлять и исправлять ошибки, таким образом корректировать учебный процесс. Использование

ЭОР также способствуют развитию навыков командной работы, навыков сотрудничества, способности обмениваться идеями и индивидуальным опытом. Что касается работы преподавателя, следует отметить важность квалифицированного сочетания возможностей и преимуществ традиционных и цифровых технологий для достижения поставленных образовательных, развивающих и воспитательных задач. Заинтересованное участие большинства студентов в мероприятиях свидетельствует о том, что Недели английского языка является настоящим интеллектуальным марафоном, переходом от трансляции знаний к интерактивному взаимодействию, действенным стимулом формирования активной жизненной позиции. Но не менее важная его задача, на наш взгляд, – создание атмосферы торжества знаний, радости познания и самосовершенствования.

ЛИТЕРАТУРА

1. «Сопровождающий электронный курс English for Economists («Английский для экономистов») по учебной дисциплине «Иностранный язык (английский)» в системе EduEnglish. bsu.by/ Ковальчук А. М. [и др.] – URL: <https://eduenglish.bsu.by/course/view.php?id=20>, регистрационный государственный номер №2.4/215 от 29.06.2022 (дата обращения: 12.06.2025)

УДК 37.016

Костенко Станислав Васильевич

аспирант кафедры романских языков
и лингводидактики

Московский Городской Педагогический
Университет

г. Москва, Россия

Stanislav Kostenko

Postgraduate student of the
Department of Romance Languages
and Linguodidactics

Moscow City Pedagogical
University

Moscow, Russia

kostenkosv105@mgpu.ru

ОСОБЕННОСТИ ПРИМЕНЕНИЯ МНЕМОТЕХНИК ДЛЯ РАЗВИТИЯ ЛЕКСИЧЕСКИХ НАВЫКОВ В УСЛОВИЯХ ЦИФРОВОЙ СРЕДЫ

В статье рассматривается влияние цифровизации на процесс обучения иностранным языкам, в частности на развитие лексических навыков. Автор отмечает, что значительная часть времени современных детей и подростков проходит в цифровой среде, что требует адаптации методик преподавания к новым условиям. В центре внимания – мнемотехники как эффективный инструмент запоминания слов и словосочетаний, способный органично интегрироваться в цифровые образовательные технологии. Приводится классификация мнемотехник по F. S. Bellezza с примерами применения, включая фонетическое кодирование, визуализацию и организационные методы. Подробно описаны варианты их реализации в цифровой среде: использование ИИ для генерации ассоциаций и изображений, интеграция в сервисы Anki, Holst, Vocabulary.com, а также игровые платформы вроде Roblox. Отмечается, что подобные инструменты позволяют не только ускорить и упростить процесс запоминания, но и повысить мотивацию обучающихся за счёт интерактивности и персонализации. Автор делает вывод о необходимости творческого подхода к сочетанию традиционных мнемотехнических приёмов с современными цифровыми ресурсами, что открывает новые перспективы для эффективного формирования устойчивых речевых навыков у школьников. Материал адресован педагогам, методистам и исследователям, работающим в сфере цифрового обучения иностранным языкам.

Ключевые слова: иностранный язык; мнемотехника; образование; цифровая среда; лексические навыки; цифровизация.

FEATURES OF THE USE OF MNEMONIC DEVICES FOR THE DEVELOPMENT OF LEXICAL SKILLS IN A DIGITAL ENVIRONMENT

The article addresses the impact of digitalization on foreign language learning, focusing on the development of lexical skills. The author emphasizes that children and adolescents spend a significant portion of their daily lives in digital environments, which requires adapting teaching methods to these conditions. Mnemonic devices are presented as effective tools for vocabulary retention, capable of integration into various digital platforms. Drawing on F. S. Bellezza's classification, the paper highlights phonetic encoding, visualization, and organizational techniques as the most relevant for lexical acquisition. Practical implementations include AI-based generation of associations and imagery, use of applications such as Anki,

Holst, and Vocabulary.com, and gamified environments like Roblox. These approaches enhance memorization efficiency and student motivation through interactivity and personalization. The study concludes that combining traditional mnemonic strategies with modern digital resources offers promising opportunities for sustainable vocabulary development in contemporary language education.

Key words: foreign language; mnemonic devices; education; digital environment; lexical skills; digitalization.

В настоящее время отечественная педагогика вынуждена отвечать на актуальные вызовы современного мира, среди которых особо остро стоит проблема цифровизации. Под цифровизацией как актуальной проблемой мы понимаем повсеместное распространение информационных технологий, прежде всего интернета и социальных сетей, во всех сферах жизни человека, в частности, в образовании [1].

Среда оказывает значительное воздействие на онтогенез человека, особенно в детском и подростковом возрасте. В настоящий момент дети и подростки по всему миру, в том числе и в России, проводят все больше времени онлайн. Фабио Дюарте приводит следующую статистику: в среднем люди проводят 6 часов 40 минут в интернете в день, а молодые люди, представители поколения Z, почти 9 часов в день [2]. Примечательно, что почти половина сегодняшних тодлеров (дети от 1 года до 3 лет) уже погружена в цифровую среду.

Мы можем констатировать, что примерно половину времени от общего времени бодрствования сегодняшние дети проводят за экранами смартфонов. А если вычесть из этого количество времени, отводимое на необходимые гигиенические процедуры, бытовые занятия и прочее, то получится даже больше половины времени бодрствования. Таким образом, среда, окружающая сегодняшних детей, преимущественно цифровая, и это обстоятельство неизбежно влияет на их развитие, образ жизни и особенности обучения. Нам, педагогическому сообществу учителей и исследователей, необходимо выявить эти особенности и адаптировать процесс обучения в соответствии с ними.

В данной статье будут рассмотрены и предложены некоторые особенности и приемы для развития лексических навыков на иностранном языке в условиях цифровой среды. Прежде всего из-за того, что именно в области обучения иностранного языка на данный момент существует великое число цифровых инструментов по сравнению с другими школьными предметами. А также из-за опыта преподавания автора статьи в данной дисциплине. Фокус на лексических навыках был сделан ввиду того, что для их развития необходимо запоминания лексических единиц – слов и словосочетаний.

Мнемотехники как эффективное средство развития лексических навыков

Одним из самых древних и незаслуженно отодвинутых на задний план средств для эффективного запоминания информации является мнемотехника. Согласно С. Д. Смирнову, «мнемотехника – это совокупность приёмов и

методов запоминания информации, основанных на использовании ассоциаций, образов, логических связей и других психических процессов для повышения эффективности запоминания и воспроизведения информации» [3]. Необходимо заметить, что помимо различной прикладной и учебной информации использование мнемотехники оправдано и распространено для запоминания иностранных слов [4].

Примечательно, что несмотря на доказанную эффективность и существующий спрос со стороны учителей, на сегодняшний день почти не существует готовых методических решений с использованием мнемотехник для их применения в образовательном процессе [5].

Рассмотрим одну из классификаций мнемотехник [6], представленной в таблице с нашими примерами.

Виды мнемотехник по Belezza F. S.

Тип мнемотехники	Метод	Описание	Пример
Организационные мнемотехники	Peg-type mnemonics (Системы с опорными точками)	Используют заранее запомненные структуры (например, слова или места) для организации информации	Метод локусов (Method of Loci): связывание элементов списка с местами в знакомом месте. Peg-word mnemonic: использование слов, ассоциирующихся с числами (например, "one is bun", "two is shoe")
	Chain-type mnemonics (Цепочные мнемотехники)	Связывают элементы списка попарно через образы или истории, создавая цепочку ассоциаций.	Link mnemonic: создание визуальных образов между парами слов. Story mnemonic: включение всех слов списка в историю.
Единичное применение (Single-use mnemonics)	Акронимы и акrostихи	Создаются специально для одного набора данных и не используются повторно.	– Первые буквы слов формируют новое слово (например, «Каждый охотник желает знать...» для цветов радуги). – Рифмы и стихи для запоминания чисел или дат.
	Использование первых букв	Опора на начальные буквы слов, чтобы создать мнемонический ключ.	СЮЗВ – «Серова Юля Завтракает Вечером». Акростих для запоминания сторон света: Север, Юг, Запад, Восток.

Кодирующие мнемотехники	Визуализация и образы	Преобразование абстрактной информации в визуальные образы.	Различные схемы, диаграммы, таблицы, mind-maps
	Фонетическое кодирование	Преобразование слов или чисел в другие слова или фразы на основе звучания или использование сходств в звучании	Caught ‘поймал’ – похоже на «кот»
	Цифрово-согласное кодирование	Присваивание цифрам буквенное обозначение (Например, 1 – к (кол), 2 – л (лебедь), 3 – т (далее по первой букве), 4 – ч и т.д.	Год крещения Руси – 988 (<i>Дал Веру Владимир</i>)

Как видно из Таблицы, для запоминания лексических единиц релевантно использовать следующие виды мнемотехник:

- Фонетическое кодирование
- Визуализация
- Организационные

Рассмотрим некоторые самые распространенные цифровые средства, которые используются для обучения иностранному языку и развитию лексических навыков, в частности. А также некоторые неочевидные решения. И предложим варианты по интеграции в них мнемотехнических приемов.

Мнемотехники в цифровой среде

III ассоциации

Зачастую не удастся придумать ассоциацию ко всем изучаемым словам, что вполне естественно. В такой ситуации чат-боты, такие как ChatGPT, Qwen или отечественный Giga Chat позволяют придумать фонетическую ассоциацию к запоминаемому слову. Заметим, что не всегда полученные результаты адекватно, но даже при неудовлетворительном результате велик шанс, что при анализе результатов появится собственная неочевидная идея.

При подготовке к урокам учитель может уточнить промпт ('запрос'), конкретизировав ассоциации к конкретному возрасту или увлечениям обучающихся.

III картинки

Говоря о чат-ботах, мы предлагаем использовать максимально их функционал, а именно способность генерировать изображения. Релевантные сервисы: Raphael AI, Midjourney или отечественный Kandinsky.

Когда ассоциация к запоминаемому слову готова, необходимо ее визуализировать для учеников. Данные сервисы позволяют сгенерировать самые необычные ассоциации за считанные секунды. Заметим, что сервисы позволяют сделать полученные изображения более индивидуальными. Например, выбрав стиль иллюстрации от реалистичной до аниме стиля.

Anki

Весьма распространенное приложение для запоминания слов через карточки можно также использовать вместе с мнемотехниками. Суть сервиса заключается в том, что пользователь создает виртуальную карточку с двумя сторонами. На одной стороне изучаемого иностранное слово, на другой – перевод. Мы предлагаем вместо перевода добавить на карточку ассоциацию. Например, сгенерированную картинку.

Онлайн доска Holst

Помимо множества полезных функций, которые можно использовать для подготовки урока по новым лексическим единицам, сервис также позволяет создавать mind maps ('ментальные карты'). С их помощью эффективно запоминать списки слов: тематические подборки или слова-исключения. Данный метод релевантно комбинировать с вышеописанными.

Vocabulary.com

Функциональный сервис, который помимо возможностей словаря предлагает и геймифицированное запоминание слов. Во-первых, мы советуем использовать его для получения более широкого контекста, примеров использования необходимого для лучшего понимания слова, а также для более эффективной ассоциации. Во-вторых, к дефинициям запоминаемых слов можно добавлять свои комментарии, которые могут быть мнемотехниками.

Roblox

Популярная игровая платформа, предоставляющая возможность самостоятельно создавать собственные игровые уровни (на сайте они называются place 'место'). Ранее мы с помощью данной платформы смогли реализовать 3 мнемотехнических приема: метод мест, где нужно было в локации «школа» расставлять кубики с английскими словами на них; метод выдуманных слов, где игрокам нужно было из кубиков с суффиксами составить выдуманное слово; метод рифм.

Таким образом, интеграция мнемотехник в цифровую образовательную среду открывает новые возможности для развития лексических навыков при изучении иностранных языков. Использование ИИ-инструментов, игровых платформ и специализированных сервисов позволяет сделать процесс запоминания слов более эффективным, интерактивным и адаптированным под индивидуальные потребности учащихся. Педагогам важно не только осваивать эти технологии, но и творчески комбинировать их с проверенными мнемоническими приемами, чтобы максимально использовать потенциал цифровизации в обучении. Внедрение таких методов в педагогическую практику способствует формированию устойчивых речевых навыков и повышению мотивации современных школьников.

ЛИТЕРАТУРА

1. Alenezi M. Digital learning and digital institution in higher education // Education Sciences. 2023. T. 13. № 1. С. 88.

2. Alarming Average Screen Time Statistics (2025) by Fabio Duarte. URL: <https://explodingtopics.com/blog/screen-time-stats> (дата обращения: 01.07.2025).
3. Смирнов С. Д. Память и мышление: учебное пособие. Аспект Пресс, 2001. 480 с.
4. Павлюкевич Л. В., Овчинникова Ю. Н. Эффективность приемов мнемотехники при формировании лексических навыков учащихся на уроке иностранного языка // Проблемы романо-германской филологии, педагогики и методики преподавания иностранных языков. 2022. № 18. С. 254–258. EDN YZWESI.
5. Костенко С. В., Мыскин С. В. Проблемы применения и методы мнемотехники в обучении иностранному языку // Экономические и социально-гуманитарные исследования. 2024. № 4 (44). С. 186–193. DOI: <https://doi.org/10.24151/2409-1073-2024-4-186-193> EDN: GRITQC.
6. Bellezza F. S. Mnemonic devices: Classification, characteristics, and criteria // Review of Educational Research. 1981. Т. 51. №. 2. С. 247–275.

УДК 81'33

Кошкарлова Наталья Николаевна
доктор филологических наук,
профессор кафедры
«Международные отношения,
политология и регионоведение»
Южно-Уральский государственный
университет
г. Челябинск, Россия

Natalya Koshkarova
Doctor of Philology, Professor
of the International Relations,
Political Science and Regional Studies
Department
South Ural State University
Chelyabinsk, Russia
studiumjuvenis@mail.ru

КОРПУСНЫЕ ТЕХНОЛОГИИ В ПРЕПОДАВАНИИ РКИ В СТРАНАХ БРИКС¹

Статья посвящена использованию корпусных технологий в преподавании русского языка как иностранного. Описаны преимущества корпусных технологий, достоинства корпусных методик, их востребованность в современных геополитических условиях и обстоятельствах. Представлены учебники и учебные пособия, которые используются при обучении русскому языку студентов из некоторых стран, входящих в БРИКС: Китай, Индия, Бразилия, ЮАР. Расширение объединения требует учета культурно-цивилизационных и языковых характеристик обучающихся, что может быть осуществлено в процессе использования национальных и диалектных языковых корпусов. Предложен алгоритм использования Национального корпуса русского языка при изучении лексики. Делается вывод о том, что использование корпусных технологий дает возможность преодолеть разрыв между теорией (практикой) преподавания русского языка как иностранного и современной международной ситуацией, решить проблему морального устаревания учебников, а в конечном счете сформировать положительный образ России как страны изучаемого языка.

Ключевые слова: корпусные технологии; русский язык как иностранный; БРИКС; образ России; культура.

CORPUS TECHNOLOGIES IN TEACHING RUSSIAN AS A FOREIGN LANGUAGE IN BRICS COUNTRIES²

The article deals with the usage of corpus technologies in teaching Russian as a foreign language. The author describes the advantages of using corpus technologies, the benefits of corpus methods, their relevance in the modern geopolitical situation. The textbooks which can be used in the process of teaching the students from the BRICS countries (China, India, Brazil, South Africa) are presented. The expansion of the BRICS organization prompts to take into account the cultural, civilization, and language peculiarities of the students who represent the BRICS countries. The national and dialect language corpora can be of great demand in this respect. The algorithm of using the Russian National Corpus in the process of teaching the

¹ Исследование выполнено за счет гранта Российского научного фонда № 24-18-00049, <https://rscf.ru/project/24-18-00049/>.

² The research is financially supported by Russian Science Foundation grant № 24-18-00049, <https://rscf.ru/project/24-18-00049/>.

word-stock is suggested. The conclusion is made that the usage of the corpus technologies makes it possible to overcome the gap between the theory (practice) of teaching Russian as a foreign language and the modern political situation, to solve the problem of the textbooks' functional depreciation and to form Russia's positive image.

Key words: corpus technologies; Russian as a foreign language; BRICS; Russia's image; culture.

С 6 по 7 июля 2025 г. в Рио-де-Жанейро (Бразилия) проходил XVII саммит БРИКС – организации, в которую в настоящее время входят не только страны-члены (Бразилия, Россия, Индия, Китай, Южная Африка, Аргентина, Египет, Иран, Объединенные Арабские Эмираты (ОАЭ), Саудовская Аравия, Эфиопия), но и страны-партнеры объединения (Белоруссия, Боливия, Индонезия, Казахстан, Куба, Малайзия, Таиланд, Уганда и Узбекистан). Бразилия председательствует в БРИКС в третий раз (ранее эта роль выпадала стране в 2010 г. и 2014 г.), а среди главных приоритетов своего председательства южно-американское государство выделяет следующие аспекты: глобальное здравоохранение; экономика и финансы; климатическая повестка; применение искусственного интеллекта; вопросы международного порядка; институциональное развитие объединения. Как видно из приведенного списка, не последнее место в череде обсуждаемых проблем отводится цифровой трансформации общества, что особенно актуально в образовательной сфере.

Не вызывает сомнения тот факт, что применение искусственного интеллекта в образовании имеет как положительные, так и отрицательные стороны. Среди плюсов внедрения цифровых технологий в преподавании ряда дисциплин (в частности, русского языка как иностранного) назовем следующие. Во-первых, адаптация возможностей образовательной среды под нужды обучающихся, которые принадлежат к так называемому цифровому поколению (*digital natives*). Во-вторых, как следствие интуитивного владения молодым поколением цифровых технологий – эволюция роли учителя, который перестает быть только транслятором знаний, а выступает в роли наставника и проводника в мире информации. В-третьих, в силу консервативного характера системы образования инновации приходят в эту сферу не в уязвимом, а в апробированном виде, что делает их более индивидуализированными, персонализированными и адаптивными под нужды конкретного обучающегося.

Одной из разновидностей цифровых технологий является корпусная лингводидактика, развивающаяся в настоящее время в двух направлениях. Первое направление (*corpus-informed*) связано с информированием составителей учебников и учебных пособий о методическом потенциале национальных языковых корпусов. Второе направление (*at the classroom door*) разрабатывает формы аудиторной работы на основе корпусных данных в зависимости от потребностей преподавателя, необходимого объема корпусных заданий, изучения того или иного языкового материала, когнитивных

особенностей студентов и многих других факторов [об этом см., например, 1]. Не вызывает сомнения тот факт, что применение корпусных технологий в практике преподавания иностранных языков расширяется в геометрической прогрессии. Лингвистические корпуса могут быть использованы не только в лингвистических целях (изучение национальной специфики аксиологически маркированных феноменов), но и в методике преподавания иностранных языков (в частности, русского как иностранного).

К достоинствам корпусных методов в ходе изучения неродного языка относятся: использование аутентичных материалов; реальные примеры использования разноуровневых языковых единиц; богатство текстового материала, который может быть использован для составления упражнений [об этом см., например, 2]. В. З. Демьянков отмечает, что корпусные данные помогают преодолеть дефицит наблюдения за использованием языка в реальной коммуникации, в том или ином контексте [3]. В своих предыдущих работах мы уже обращались к проблеме использования корпусных технологий при изучении английского языка [4] и русского языка как иностранного [5, 6]. В современных геополитических условиях и обстоятельствах востребованность корпусных технологий возрастает в связи с расширением популярности русского языка, в частности, в государствах-членах БРИКС.

Так, в марте 2025 г. международная сеть TV BRICS представила новый образовательный проект «Уроки русского с Таней Сёмке», который направлен на изучение русского языка и русской культуры. В 2024 новые образовательные платформы по изучению русского языка были созданы в Бразилии, Объединенных Арабских Эмиратах и Саудовской Аравии. Как видно из приведенного списка мероприятий по продвижению русского языка в мировом образовательном пространстве, такая работа проводится в странах с разными культурными особенностями, а целевой аудиторией в процессе обучения являются студенты, для которых родным являются типологически различные языки. Корпусные технологии помогают нивелировать указанные различия – культурный и языковой фактор – и сделать процесс изучения неродного языка наиболее интересным и плодотворным.

В отечественной и зарубежной методике преподавания РКИ накоплен значительный опыт обучения студентов из Китая, Индии, португало- и англоговорящих студентов. Так, для работы с китайскими студентами давно и успешно используется УМК «Дорога в Россию» [7]. В 2016 г. вышел в свет учебник интерактивный учебник по русскому языку для китайцев [8]. Для работы с индийскими учащимися применяются как учебники, изданные непосредственно для студентов из Индии, так и учебники для англоговорящих слушателей. Учебники для студентов, говорящих на английском языке, могут с успехом использоваться в обучении выходцев из ЮАР: Интенсивный курс русского языка. Месяц в Москве [9]; Учебник русского языка для лиц, говорящих на английском языке [10]; Учебник русского языка

для индийцев через посредство английского языка [11]; New (Audio-Enabled) Essential Russian [12]; Rush into Russian: Basic Russian with Illustrations [13]; Russian for Indians [14]. Для португалоговорящих студентов используется широкий спектр изданий: начиная от лексикографических источников и заканчивая специализированными учебными пособиями, предназначенными для обучения студентов определенной специальности. Из известных нам учебников перечислим следующие: Учебный русско-португальский словарь [15]; Самоучитель русского языка с упражнениями, отрывками для чтения и словарем [16]; Речевой этикет. Русско-португальские соответствия: справочник [17]; Учебное пособие по практическому курсу военного перевода [18]; Первое путешествие в мир русских слов (пособие лексикографического типа для носителей португальского языка) [19]; Dialog – Manual de Russo como Língua Estrangeira para Falantes de Português, nível A1 [20].

Как видно из приведенного списка, учебно-методическое обеспечение процесса преподавания русского языка как иностранного достаточно разнообразно и позволяет учитывать как языковые особенности студентов в процессе изучения РКИ, так и культурно-цивилизационные различия стран, которые они представляют. Однако по мере расширения БРИКС и присоединения новых стран преподаватели РКИ ощущают дефицит учебников, которые бы отражали типологические и лингвокультурологические характеристики таких стран как Египет, Индонезия, Эфиопия и др. Восполнить эту лакуну и призваны корпусные технологии, которые могут дать представление как о нормативном, так и диалектном варианте использования русского языка.

Для этих целей в Национальном корпусе русского языка (НКРЯ) может с успехом применяться следующие возможности: Обучающий корпус [21], Портрет слова [22], упражнения из НКРЯ [23]. На более продвинутом этапе изучения русского языка как иностранного возможно обращение к диалектным корпусам [24], которые отражают «понимание значимости местных народных говоров как первоосновы национального языка и национальной ментальности» [25, с. 55]. Применение как Национального корпуса русского языка, так и диалектных корпусов познакомит с лексикой и грамматикой русского языка, сформирует представление об истории и культуре страны изучаемого языка, повысит функциональную и цифровую грамотность студентов.

Несмотря на возможные временные и ресурсные затраты при работе с корпусом, корпусные технологии позволяют расширить цифровые возможности и самих преподавателей РКИ. Одновременно с этим решается задача морального устаревания учебников и учебных пособий по РКИ, когда они отражают не всегда актуальный материал, что в дальнейшем приводит к формированию ложного и даже отрицательного образа России. Например, НКРЯ может использоваться при изучении лексики. Примерный алгоритм использования лингвистического корпуса в данном случае может быть основан не только на использовании основного корпуса, но и различных

подкорпусов в зависимости от цели и задачи урока, культурных особенностей студентов, их уровня владения русским языком и других факторов. Так, востребованными могут оказаться упражнения на сочетаемость глаголов, изучение приставочных глаголов, то есть тем, которые представляют наибольшую сложность при изучении русского языка. Корпусные онлайн-ресурсы – это разновидность информационно-коммуникационных технологий, внедрение которых отвечает духу времени и потребностям целевой аудитории

Таким образом, использование корпусных технологий в современной геополитической ситуации позволяет учитывать культурно-цивилизационные различия студентов, представляющих страны, входящие в объединение БРИКС. В настоящее время методика преподавания РКИ пока отстаёт от соответствующей области при изучении английского языка в плане включения в тексты и задания учебников актуальных материалов по современным вопросам международного сотрудничества. Преодолеть разрыв между теорией (практикой) РКИ и политическими реалиями призваны корпусные технологии, которые позволяют реализовывать основные принципы коммуникативной методики, когда акцент делается не на механическом заучивании слов и грамматических правил, а решается задача предоставления студентам инструментов общения, которые могут с успехом применяться в реальных бытовых и профессиональных ситуациях. Корпусные технологии в данном случае выполняют и важную социальную миссию – формирование позитивного образа России как уникальной страны-цивилизации с богатыми историческими, культурными и языковыми традициями, которые в силу объективных факторов не всегда могут быть корректно представлены в учебниках, но отражены в национальном и диалектном корпусах.

ЛИТЕРАТУРА

1. Алексеев В. А. Алексеева Н. И., Румянцева Н. М. Интенсивный курс русского языка. Месяц в Москве: Учеб. пособие. М.: Изд-во РУДН, 1994. 160 с.
2. Балыхина Т. М., Румянцева Н. М. Русский язык: интерактивный учебник по русскому языку для китайцев. А1, А2, В1 уровни. М. : РУДН, 2016. 455 с.
3. Вагнер В. Н., Овсиенко Ю. Г. Учебник русского языка для лиц, говорящих на английском языке. New Delhi: People's Publishing House Private Limited, 2023. 502 с.
4. Вострякова Н. А. Первое путешествие в мир русских слов (пособие лексикографического типа для носителей португальского языка). Волгоград: ВолгГТУ, 2017. 152 с.
5. Горина О. Г., Царакова Н. С. Корпусные инструменты, маршруты и эксперименты в современной лингводидактике // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2021. Т. 19. № 2. С. 36–53.

6. Демьянков В. З. Об усвоении неродного языка и корпусной лингвистике: что такое хайли лайкли? // Вестник Московского университета. Серия 19: Лингвистика и межкультурная коммуникация. 2021. № 2. С. 70–81.
7. Дорога в Россию: учебник русского языка / Антонова В. Е., Нахабина М. М., Сафронова М. В., Толстых А. А. Ч1. 342с. Ч2. 256 с. Ч3. 200с. Ч4. 184с.
8. Земсков М. В. Возможности использования данных корпусной лингвистики в процессе преподавания РКИ // Майские чтения (Язык и репрезентация культурных кодов): материалы XIII Всероссийской (с международным участием) научно-практической конференции молодых ученых. – Самара: Самарский национальный исследовательский университет имени академика С. П. Королева, 2024. – С. 121–124.
9. Иванцова Е. В. Томский диалектный корпус: обоснование концепции и перспективы развития // Вопросы лексикографии. 2017. № 11. С. 54–70.
10. Кетайдене Е. И., Верхуша В. М., Митрохина В. И. Учебный русско-португальский словарь. М. : Русский язык, 1983. 288 с.
11. Кошкарлова Н. Н. Составление упражнений на базе лингвистического корпуса при изучении РКИ // Когнитивные стратегии филологического образования в России и за рубежом : Сборник научных статей Всероссийской научно-практической конференции с международным участием, посвященной 130-летию со дня рождения Н. С. Трубецкого. Екатеринбург: Уральский государственный педагогический университет, 2020. С. 98–103.
12. Koshkarova N. Corpus Technologies in Teaching English to Speakers of Other Languages // *Numanities – Arts and Humanities in Progress*. 2021. Vol. 20. P. 107–123.
13. Кошкарлова Н. Н., Шарафутдинова О. И. Лингвистический корпус как средство обучения лексике на продвинутом этапе изучения русского языка как иностранного // Динамика языковых и культурных процессов в современной России. 2022. № 7. С. 1308–1313.
14. Мещеряков О. В., Ротаренко А. Д. Учебное пособие по практическому курсу военного перевода: Португальский язык. Нижний Новгород: Нижегородский университет, 2008. 318 с.
15. Обучающий корпус. URL: <https://ruscorpora.ru/search?search=CgQyAggH> (дата обращения: 10.07.2025).
16. Портрет слова. URL: <https://ruscorpora.ru/word/school> (дата обращения: 10.07.2025).
17. Самоучитель русского языка с упражнениями, отрывками для чтения и словарем / сост. Марина Доленга. Editora Globo. Rio de Janeiro – Porto Alegre – Sao Paolo, 1955. 317 с.
18. Тинякова Е. А. Учебник русского языка для индийцев через посредство английского языка. Екатеринбург: Ridero, 2025. 341 с.
19. Томский диалектный корпус. URL: <http://losl.tsu.ru/corpus> (дата обращения: 10.07.2025).

20. Упражнения. URL: <https://ruscorpora.ru/exercise> (дата обращения: 10.07.2025).
21. Формановская Н. И., Астахова Е. В. Речевой этикет. Русско-португальские соответствия: справочник. М. : Высшая школа, 2004. 109 с.
22. Maurya A, Zhilina O. New (Audio-Enabled) Essential Russian. Delhi: Goyal Publishers, 2021. 558 p.
23. Nailia Baldé, Jayanti Dutta, Ana Carina Prokopyshyn. Dialog – Manual de Russo como Língua Estrangeira para Falantes de Português, nível A1, 2ª ed., Coimbra: Grácio Editor, 2017. P. 173.
24. Pande H. Ch. Russian for Indians. Delhi: Goyal Publishers, 2023. 328 p.
25. Roitberg N. V. Rush into Russian: basic Russian with Illustrations. Course Book. Kazan: Buk, 2019. 194 p.

УДК 372.881.1

Крылова Елена Александровна

кандидат педагогических наук,
доцент Высшей школы лингвистики
и педагогики

Санкт-Петербургский политехни-
ческий университет Петра Великого
г. Санкт-Петербург, Россия

Elena Krylova

PhD in Pedagogy, Associate Professor
of the Institute of Humanites
Peter the Great St. Petersburg
Polytechnic University
Saint-Petersburg, Russia
krylova_ea@spbstu.ru

Одинокая Мария Александровна

кандидат педагогических наук,
доцент Высшей школы лингвистики
и педагогики

Санкт-Петербургский политехни-
ческий университет Петра Великого
г. Санкт-Петербург, Россия

Maria Odinskaya

PhD in Pedagogy, Associate Professor
of the Institute of Humanites
Peter the Great St. Petersburg
Polytechnic University
Saint-Petersburg, Russia
odinokaya_ma@spbstu.ru

РАЗВИТИЕ ИНОЯЗЫЧНЫХ ЛЕКСИЧЕСКИХ НАВЫКОВ С ИСПОЛЬЗОВАНИЕМ SPACED REPETITION SYSTEM И ЦИФРОВЫХ ИНСТРУМЕНТОВ

Статья посвящена исследованию применения системы интервальных повторений (Spaced Repetition System) в образовательной практике, направленной на развитие иноязычных лексических навыков. Рассматриваются ключевые характеристики системы интервальных повторений, основанные на принципах кривой забывания Г. Эббингауза, а также методический потенциал цифровых инструментов, таких как Anki, Quizlet, Memrise и других, в контексте обучения иностранным языкам в вузе. Особое внимание уделено персонализации обучения, контекстуализации материала, геймификации и автоматизации процесса повторения иноязычных лексических единиц. Представлены результаты педагогического эксперимента, в ходе которого студенты, использовавшие систему интервальных повторений, показали значительно более высокие результаты в усвоении иноязычной лексики по сравнению с контрольной группой. Исследование подтверждает, что сочетание системы интервальных повторений с цифровыми технологиями не только повышает эффективность запоминания, но и способствует формированию устойчивых лексических навыков, что особенно важно в условиях ограниченного времени аудиторных занятий в высшей школе.

Ключевые слова: лексический навык; система интервальных повторений; кривая забывания; цифровые инструменты; иностранные языки.

DEVELOPING FOREIGN LANGUAGE VOCABULARY SKILLS USING THE SPACED REPETITION SYSTEM AND DIGITAL TOOLS

The article is devoted to the study of the application of the Spaced Repetition System in educational practice aimed at developing foreign language lexical skills. It examines the key characteristics of the spaced repetition system, based on the principles of Hermann Ebbinghaus's forgetting curve, as well as the methodological value of digital tools such as Anki, Quizlet,

Memrise, and others in the context of foreign language learning in higher education. Particular attention is paid to the personalization of learning, contextualization of material, gamification, and automation of the process of repeating foreign language lexical units. The article presents the results of a pedagogical experiment, during which students who used the spaced repetition system demonstrated significantly higher results in mastering foreign language vocabulary compared to the control group. The study confirms that combining the spaced repetition system with digital technologies not only enhances memorization efficiency but also contributes to the development of stable lexical skills, which is particularly important given the limited time allocated for classroom instruction in higher education.

Key words: lexical skills; spaced repetition system; forgetting curve; digital educational tools; foreign languages.

Одним из ключевых аспектов изучения любого иностранного языка является развитие лексических навыков, которые служат фундаментом для всех видов речевой деятельности – аудирования, говорения, чтения и письма, определяя коммуникативную успешность индивида. Однако процесс усвоения иноязычной лексики нередко сопряжён с трудностями, связанными с отсутствием языковой среды; недостаточным количеством часов, выделяемых на дисциплину «Иностранный язык» в школе или вузе; а также со значительным объёмом лексического материала, необходимостью его регулярного повторения для сохранения в долговременной памяти.

В свете вышесказанного определённый интерес для практики преподавания иностранных языков представляют методы, основанные на когнитивной психологии, в частности, Spaced Repetition System (SRS) или система интервальных повторений. SRS заключается в оптимизированном повторении учебного материала через постоянно увеличивающиеся интервалы времени, что способствует переносу лексических единиц из кратковременной в долговременную память обучающихся. В сочетании с цифровыми инструментами (Anki, Quizlet, Memrise, SuperMemo и др.) SRS представляет собой перспективный инструмент для повышения эффективности изучения иноязычной лексики.

Целью данной статьи является анализ теоретических основ применения Spaced Repetition System в обучении иноязычной лексике, а также обзор и оценка возможностей современных цифровых инструментов, позволяющих реализовать данный подход в практике преподавания иностранных языков в высшей школе.

Система интервальных повторений основана на принципах «кривой забывания» Г. Эббингауза, который эмпирически доказал, что большая часть информации (до 70 % в первые сутки) теряется безвозвратно, если не предпринимать усилий по ее закреплению. При этом также было установлено, что стратегически спланированные повторения в моменты, предшествующие моменту забывания, не только активизируют память, но и существенно замедляют скорость забывания в будущем, укрепляя нейронные связи [1].

Исследованиями вопросов забывания и запоминания в области изучения иностранных языков занимался американский лингвист и психолог Пол

Пимслер, разработавший систему Graduated Interval Recall (градуированное интервальное воспоминание), в фокусе которой лежит принцип аудиоцентричности. Так, обучающийся слушает диалоги на иностранном языке, повторяет услышанное и отвечает на вопросы. При этом диалоги содержат повторяющиеся через растущие промежутки времени лексические единицы. Новая лексическая единица вводится на уроке, затем сознательно повторяется через определенные временные интервалы (5 секунд → 25 секунд → 2 минуты → 10 минут → 1 час → 5 часов → 1 день → 5 дней → 25 дней) [2].

Немецкий психолог Себастьян Лейтнер продолжил исследования проблемы забывания и запоминания и в 1970-х гг. разработал систему интервального повторения с использованием физических карточек и коробок. С. Лейтнер предложил дифференцированный подход к повторению материала в зависимости от уровня его усвоения [3]. Так, обучающийся использует от трех до пяти пронумерованных коробок, каждая из которых соответствует определенной частоте повторения лексических карточек:

- *Коробка 1* содержит новые карточки и те, которые представляют наибольшие трудности, и повторяется каждый день.
- *Коробка 2* содержит карточки с лексическими единицами, которые обучающийся знает умеренно хорошо. Работа с ними ведется через день или раз в 2–3 дня.
- *Коробка 3* содержит карточки с лексическими единицами, которые обучающийся знает хорошо. Работа с ними ведется раз в неделю.
- *Коробка 4* содержит карточки с лексическими единицами, которые обучающийся знает очень хорошо. Обучающийся повторяет такие лексические единицы один раз в две недели.

По мере овладения лексическими единицами, обучающийся перемещает карточки в следующую коробку с более редким интервалом повторения. Данная система стала фундаментом для современных цифровых инструментов, работающих на основе Spaced Repetition System.

Современные исследователи трактуют систему интервальных повторений как метод обучения, который предполагает повторение материала через увеличивающиеся интервалы времени для оптимизации запоминания и сохранения информации в долговременной памяти [4; 5]. Вопросы развития иноязычных лексических навыков с использованием системы интервальных повторений на основе цифровых инструментов представляют интерес для отечественных и зарубежных исследователей. Так, исследование Е. Крюковой подтвердило высокую эффективность метода интервального повторения на основе цифровых инструментов (Anki, Quizlet) в сравнении с традиционными методами обучения английскому языку. Студенты экспериментальной группы продемонстрировали положительную динамику в развитии лексических навыков (идентификация изученных лексических единиц, использование изученных лексических единиц в речи). 65 % студентов экспериментальной группы считают метод высокоэффективным для изучения новой лексики [6].

G. Alamelu и M. Ilangkumaran в своем исследовании изучают влияние методов интервального повторения на усвоение иноязычной профессиональ-

ной терминологии студентами инженерных специальностей [7]. Исследователи приходят к выводу, что эффективность SRS зависит от ряда факторов: формат карточки (контекст, мультимедиа), регулярность занятий, сознательное применение метакогнитивных стратегий (анализ ошибок, использование мнемоники) и т.д.

K. Mueller и G. Hays анализируют результаты внедрения приложения WordEngine в иноязычный учебный процесс. WordEngine – билингвальное приложение, которое позволяет создать персонализированный список слов для каждого студента, а также предоставляет возможность создавать визуальные, аудиальные и контекстуальные лексические задания. Авторы считают опыт использования WordEngine перспективным для развития иноязычных лексических навыков студентов: за год студенты увеличили словарный запас на 1153 слова (40 % прироста); темп изучения новых лексических единиц в 4,2 раза быстрее традиционных методов [8].

В своей статье М. И. Иголкина и В. С. Язынина изучают дидактический потенциал современных цифровых платформ, основанных на методе интервальных повторений:

- 1) Anki – создание лексических карточек с определениями и примерами.
- 2) Memrise – цифровая платформа для изучения иностранных языков с рейтинговой системой.
- 3) Tinycards – создание подборок лексических единиц по разным темам.
- 4) Quizlet – создание лексических карточек с определениями и примерами, мини-игры для запоминания лексических единиц.
- 5) Wordsteps – платформа заданиями на отработку лексических единиц по разным темам.
- 6) Lexilize flashcards – создание лексических карточек с иллюстрациями, мини-игры для запоминания лексических единиц.

Авторы приходят к выводу, что использование цифровых инструментов с методом интервальных повторений может значительно ускорить процесс обучения [9].

Louis Lafleur и Yu Kanazawa анализируют результаты внедрения разработанного программного обеспечения Interleaved Spaced Repetition Software – ISRS в иноязычный учебный процесс. Оценка уровня развития лексических навыков студентов осуществляется по трем показателям – форма, использование, значение (form, meaning, use). Экспериментальная группа, использовавшая ISRS, статистически значимо превзошла контрольную группу по общему приросту баллов. Наиболее значимым, по мнению исследователей, был прирост в категории "Использование", что позволяет сделать вывод о положительном влиянии методов интервального повторения на развитие продуктивных лексических навыков [10].

Анализ научных исследований позволил сформулировать значимые характеристики системы интервальных повторений в контексте усвоения иноязычной лексики с помощью цифровых инструментов.

1. *Персонализация процесса усвоения лексических единиц.* Обучающийся/преподаватель создает лексические карточки, соответствующие его

уровню владения иностранным языком, целям и интересам. В свою очередь алгоритмы цифровых платформ адаптируются к индивидуальной скорости усвоения лексических единиц и успешности их запоминания.

2. *Контекстуализация и мультимедийность.* Цифровые лексические карточки могут включать не только лексическую единицу и перевод, но и аудио произношения лексической единицы, изображения, примеры употребления в контексте (как в текстовом, так и видео формате), синонимы/антонимы, грамматические характеристики и т.д. То есть речь идет о полноценном формировании лексического навыка, а не о механической ассоциации «слово-перевод».

3. *Геймификация процесса усвоения лексических единиц.* Цифровые платформы используют игровые механики (очки, уровни, рейтинги, награды и т.д.), что повышает вовлеченность обучающихся и превращает рутинное повторение лексических единиц в увлекательный процесс.

4. *Автоматизация процесса усвоения лексических единиц.* Цифровые платформы рассчитывают индивидуальные повторения на основе алгоритмов, что позволяет кастомизировать процесс усвоения лексических единиц под индивидуальные нужды обучающихся.

Таким образом, система интервальных повторений оптимизирует процессы консолидации памяти и является перспективным решением проблемы прочного усвоения иноязычной лексики.

С целью внедрения системы интервальных повторений в иноязычный учебный процесс был проведен анализ цифровых инструментов SRS (Таблица 1). После изучения возможностей цифровых платформ, авторы статьи остановились на платформе «Взнания», позволяющей создавать подборки лексических единиц, разрабатывать разнообразные лексические упражнения, а также отслеживать прогресс обучающихся.

Таблица 1

Сравнительная таблица SRS-инструментов

Критерий / Инструмент	Anki	Quizlet	Memrise	SuperMemo	Взнания
Алгоритм	SM-2 (настраиваемый)	Упрощенный адаптивный	Проприетарный (умеренно сложный)	SM-18 (наиболее сложный)	Упрощенный адаптивный (система Лейтнера)
Пользовательский опыт	Сложный для новичков	Интуитивный	Игровой интерфейс	Устаревший дизайн	Интуитивный
Геймификация	Отсутствует	Игры match, gravity	Уровни, очки, мемы	Нет	Очки, рейтинги, достижения, игры

Доступность	Бесплатно (кроме iOS), оффлайн	Бесплатно, версия Pro с полным функционалом	Бесплатно, версия Pro с полным функционалом	Только подписка	Бесплатно, версия Pro с играми
Аудио	Запись/импорт TTS	Синтез речи	Видео с носителями	Да	Синтез речи
Контекст	Cloze, примеры	Только в платной версии	Предложения	Да	Контекстные примеры
Упражнения	Карточки + дополнительные плагины	Тесты, игры	Тесты, игры	Многоформатные упражнения	Карточки (обучение, запоминание) Тесты (выбор ответа, написание) Игры
Для преподавателей	Нет интеграции	Классы, отслеживание	Нет интеграции	Нет интеграции	Создание групп Назначение карточек Статистика Проверка письменных ответов

Кратко опишем алгоритм внедрения платформы «Взнания» в иноязычный учебный процесс. Педагогический эксперимент проводился в Санкт-Петербургском политехническом университете Петра Великого в течение 3 месяцев (с февраля 2025 г. по май 2025 г.). В исследовании приняли участие 40 магистрантов первого курса, обучающихся по направлению «Цифровая лингвистика» (уровень владения английским языком B1+/B2; количество занятий в неделю – 2 академических часа). Студенты были разделены на две группы по 20 человек: экспериментальную и контрольную. Обе группы обучались у одного преподавателя по учебнику Language Hub Upper-Intermediate, Macmillan Education.

Студентам обеих групп был предложен входной тест, разработанный авторами на основе отобранных из учебника лексических единиц (Уроки 5–8). Тест включал задания на подстановку, соотнесение, выбор правильного варианта ответа, ряд открытых вопросов. Максимальное количество баллов – 50. Результаты входного тестирования представлены в Таблице 2.

Таблица 2

Результаты входного тестирования

Группа	Количество студентов	Среднее	Стандартное отклонение
ЭГ	20	20,7	1.813
КГ	20	20.4	1.080

Аудиторный учебный процесс в обеих группах был примерно одинаковым и обеспечивал комплексное развитие иноязычной коммуникативной компетенции студентов. Внеаудиторная работа студентов контрольной группы была организована традиционным образом, без использования цифровых инструментов и SRS. Студенты выполняли домашние задания, нацеленные на закрепление лексических единиц (заполнить пропуски, соотнести слово с его определением, кроссворд, перевести предложения на английский и т.д.).

Студенты экспериментальной группы выполняли задания на платформе Взгляни в рамках внеаудиторной работы. По каждому уроку учебника были отобраны лексические единицы, на основе которых были разработаны лексические карточки и интерактивные лексические задания на платформе Взгляни. После введения и первичной отработки новых лексических единиц на занятии студентам предлагалась работа с лексическими карточками на платформе. Отметим, что платформа позволяет реализовать разные форматы работы с лексическими карточками: соотнести ЛЕ с определением, послушай и выбери правильный вариант ответа, прочитай определение и напиши слово и т.д., что позволяет сформировать прочную основу для дальнейшего развития продуктивного лексического навыка (Рисунок 1).

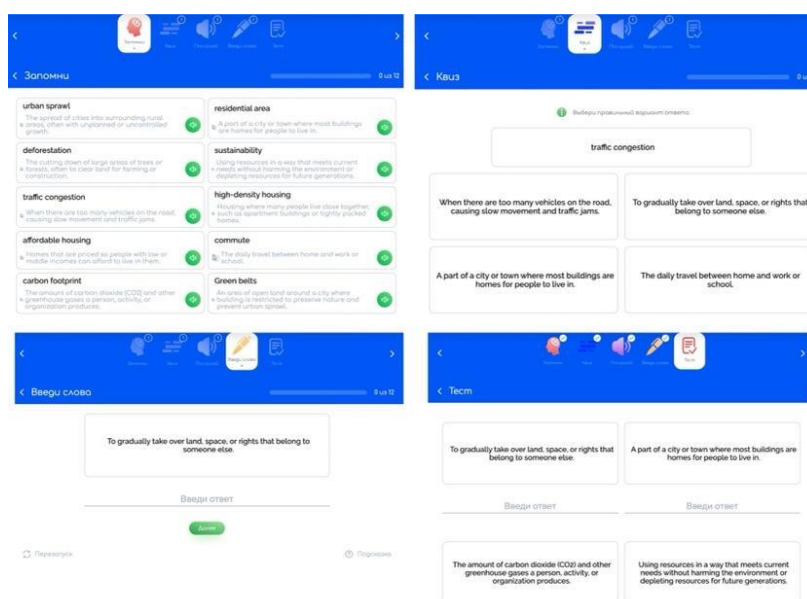


Рис 1. Примеры заданий с лексическими карточками

Далее через определенные промежутки студентам назначались интерактивные упражнения для дальнейшего закрепления лексических единиц в контексте (Рисунок 2):

- День 1. Первичное заучивание лексических единиц на занятии.
- День 2. Повторение. Работа с лексическими карточками на платформе.
- День 4. Повторение. Контекстное интерактивное упражнение на платформе.
- День 7. Повторение. Контекстное интерактивное упражнение на платформе.
- День 14. Повторение. Коммуникативное упражнение на платформе.
- День 21. Контрольное повторение. Тест на платформе.

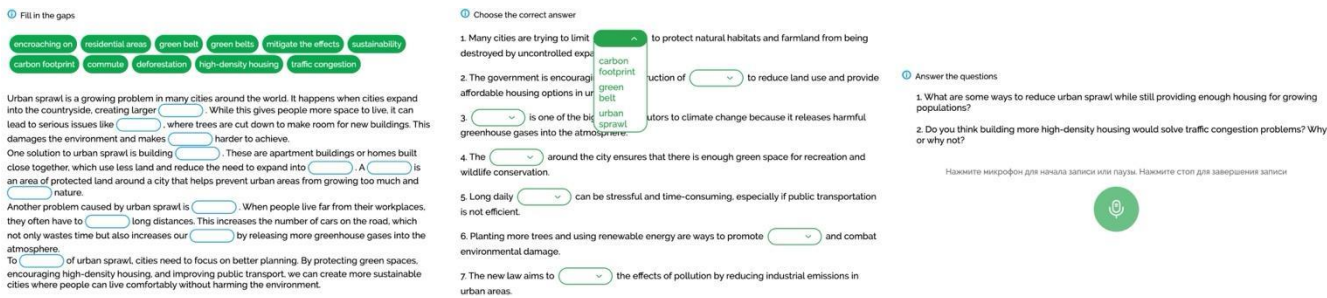


Рис 2. Примеры интерактивных лексических заданий

Далее был проведен выходной лексический тест, также разработанный авторами на основе отобранных из учебника лексических единиц (Уроки 5-8). Максимальное количество баллов составляло 50. Результаты итогового тестирования представлены в Таблице 3.

Таблица 3

Результаты итогового тестирования

Группа	Количество студентов	Среднее	Стандартное отклонение
ЭГ	20	38,2	3.134
КГ	20	30,4	2.331

Таким образом, сравнение результатов входного и выходного тестирования показало улучшение лексических навыков у студентов экспериментальной группы и незначительный прирост в контрольной. Полученные результаты подтверждают, что внедрение системы интервального повторения в сочетании с современными технологиями активизируют процесс развития лексических навыков студентов. Spaced Repetition System позволяет оптимизировать процесс усвоения иноязычных лексических единиц и их

сохранения в долговременной памяти студента, а цифровые инструменты предоставляют широкий спектр возможностей для интерактивного и гибкого обучения.

ЛИТЕРАТУРА

1. Ebbinghaus H. Memory: A contribution to experimental psychology. Columbia University, 1913. 123 p.
2. Pimsleur P. A memory schedule // The Modern Language Journal. 1967. № 2. P. 73–75.
3. Latanova R. U., Aubakirova S. S. SRL as functional approach in educational activity // Functional aspects of intercultural communication. Translation and interpreting issues: Collection of articles. Vol. 7. Moscow: Российский университет дружбы народов (РУДН). 2020. P. 568–574. DOI: <https://doi.org/10.22363/2686-8199-2020-7-568-574>.
4. Tabibian B., Upadhyay U., De A. Enhancing human learning via spaced repetition optimization. Proceedings of the National Academy of Sciences. 116(10). 2019. P. 3988–3993. DOI: <https://doi.org/10.1073/pnas.1815156116>
5. Jacinto A. S., de Medeiros F. P. A., de Sousa M. P. An Approach to Spaced Repetition Methodology for Teaching Markup and Scripting Programming Languages. 2024 IEEE Frontiers in Education Conference (FIE). 2024. P. 1–9. DOI: <https://doi.org/10.1109/fie61694.2024.10893444>
6. Kryukova Y., Yamshynska N., Kutsenok N. Enhancing ELS students' language proficiency through Spaced Repetition // Education. Innovation. Practice. 2024. №12(8). P. 34–41. DOI: <https://doi.org/10.31110/2616-650x-vol12i8-005>
7. Alamelu G., Ilankumaran M. Factors affecting consciously improving vocabulary with a spaced repetition system // International Journal of System of Systems Engineering. 2024. 14(3). P. 260–277. DOI: <https://doi.org/10.1504/ijssse.2024.138349>.
8. Mueller K., Hays G. Vocabulary learning outside of the classroom: Institutional use of a spaced repetition // Vocabulary Learning and Instruction. 2022. №11(2). P. 48–56. DOI: <https://doi.org/10.7820/vli.v11.2.mueller>
9. Иголкина М. И., Язынина В. С. Метод интервальных повторений при обучении иностранным языкам // Гуманитарный вестник. 2023. №3 (101).
10. Lafleur L., Kanazawa Yu. The Effects of Interleaved Spaced Repetition Learning on Vocabulary Knowledge // CALL-EJ. 2024. №25. P. 172–194.

УДК [004.8:378.016]:[811.161.1:81'243]

Лебедева Диана Валерьевна
магистр педагогических наук,
старший преподаватель кафедры
белорусской и зарубежной
филологии
Белорусский государственный
университет культуры и искусств
г. Минск, Беларусь

Dziyana Lebedzeva
Master of Pedagogical Sciences,
Senior Lecturer at the Department
of Belarusian and Foreign Philology
Belarusian State University
of Culture and Arts
Minsk, Belarus
dlebedeva@yahoo.com

АУТЕНТИЧНОСТЬ СГЕНЕРИРОВАННЫХ УЧЕБНЫХ ТЕКСТОВ В ОБУЧЕНИИ ИНОСТРАННОМУ ЯЗЫКУ

Сгенерированные ИИ тексты проникают на уроки иностранного языка через пособия и тексты, созданные как носителями языка, так и преподавателями для нужд конкретных групп учащихся. Целью нашего исследования является выяснить, обладают ли сгенерированные тексты качеством аутентичности, как ее понимают в методике обучения иностранному языку. Аутентичными могут считаться тексты, заимствованные из коммуникативной практики носителей языка и изначально не предназначенные для изучения иностранного языка. Они позволяют учащимся иметь непосредственный доступ к языку, культуре и коммуникации носителей языка. Исходя из данного выше определения, тексты, сгенерированные ИИ не для обучения, а для реальных ситуаций общения в стране изучаемого языка, можно считать аутентичными. Если же текст был сгенерирован или трансформирован с помощью ИИ для образовательных целей, т.е. с учетом потребностей учащихся (тематики, объема, стиля, уровня владения языком, необходимости отработки определённых лексических или грамматических конструкций), он не является аутентичным. Особую проблему в работе со сгенерированными текстами представляет неточность содержащейся в них информации.

Ключевые слова: изучение иностранного языка; учебный языковой текст; аутентичный текст; аутентичность; генеративные языковые модели; искусственный интеллект.

AUTHENTICITY OF AI-GENERATED EDUCATIONAL TEXTS IN FOREIGN LANGUAGE TEACHING

AI-generated texts are increasingly entering foreign language lessons through textbooks and texts created both by native speakers and teachers for the needs of specific groups of learners. The aim of our study is to determine whether AI-generated texts possess the quality of authenticity as understood in foreign language teaching methodology. Authentic texts are those borrowed from the communicative practice of native speakers and originally not intended for foreign language learning. They allow learners to have direct access to the language, culture, and communication of native speakers. Based on this definition, texts generated by AI not for educational purposes but for real communication situations within the target language country can be considered authentic. However, if a text has been generated or transformed by AI specifically for educational purposes – i.e., taking into account learners' needs (topics, length,

style, proficiency level, necessity to practice certain lexical or grammatical constructions) – it is not authentic. A particular challenge when working with AI-generated texts is the inaccuracy of the information they may contain.

Key words: foreign language learning; educational language text; authentic text; authenticity; generative language models; artificial intelligence.

Информационные технологии стали неотъемлемой частью современной жизни. Искусственный интеллект (ИИ), подтверждая прогнозы ученых, уже начал заменять собой человека в некоторых не только рабочих, но и в творческих профессиях. Такой прогресс в общении машин и людей был достигнут благодаря технологии обработки естественного языка (Natural language processing, NLP). Она же лежит в основе высокотехнологичных вещей, которыми мы пользуемся каждый день: от виртуальных помощников вроде Сири и Алисы до инструментов языкового перевода и интеллектуального набора текста (T9). Технология NLP, по сути, позволяет компьютерам понимать людей и говорить, как они. Генеративные языковые модели, использующие NLP, в состоянии улавливать логику человеческой речи (письменной и устной) и создавать собственные тексты по запросам пользователя. Нейросети создают развернутые ответы на поисковые запросы, пишут новости, обзоры, передают краткое содержание сайтов, фильмов, книг, ведут в чат-ботах и приложениях диалоги с пользователями, имитируя реального собеседника.

Компьютер обрел голос, стал участником речевого акта, автором, творящим при этом на множестве языков и диалектов. Информационное пространство вокруг нас заполняется текстами, искусственно сгенерированными машинами, которые с каждым днем все сложнее отличить от текстов, созданных людьми. Каким образом созданные машинами тексты повлияют на язык и письменность людей в долгосрочной перспективе, пока говорить рано, но уже сегодня нужно признать, что сгенерированные ИИ тексты влияют на процесс обучения родному и иностранным языкам.

Целью обучения иностранным языкам является формирование иноязычной коммуникативной компетенции, позволяющей использовать иностранный язык как средство межличностного, межкультурного и профессионального взаимодействия в повседневной, социально-культурной и трудовой сферах. Для достижения этой цели требуются соответствующие инструменты обучения. Среди них традиционно присутствуют педагогические подходы, методы и приемы, но в основе обучения всегда лежит текст, как лексический, грамматический, фонетический и стилистический образец употребления языка его носителями в устной или письменной форме. Текст в контексте методики – это вербальное произведение, напечатанное или бытующее в устной форме, начиная с произведений литературы и фольклора и заканчивая повседневными диалогами и списком покупок. Текст всегда относительно завершен по своему содержанию и строению, имеет автора и адресата.

Именно анализ доступных иноязычных текстов позволяет педагогам отобрать лексический и грамматический материал для обучения иностранному языку. Изучая заимствованные иноязычные тексты, учащиеся в итоге приходят к созданию собственных текстов на иностранном языке, т.е. к говорению и письму. Таким образом, уровень достигнутой иноязычной коммуникативной компетенции будет в значительной степени зависеть от качества используемых в обучении текстов.

Каким образом сгенерированные ИИ тексты становятся материалом для изучения иностранного языка? Есть пять основных путей, как иноязычные тексты становятся частью обучения: 1) они входят в состав учебного пособия, созданного носителями языка; 2) включены в учебное пособие, созданное не носителями языка в стране обучения; 3) сочиняются самими преподавателями, исходя из целей конкретного урока и особенностей обучающихся; 4) подбираются преподавателями из иноязычных источников (доступной литературы и Интернета); 5) генерируются преподавателями с помощью ИИ, исходя из нужд обучения. Здесь мы сталкиваемся со сложностью определения происхождения текстов: были ли они сгенерированы ИИ или нет. Если нас интересуют только искусственно созданные тексты, то к ним явно относятся тексты категории 5, совсем не относятся тексты категории 3, и они могут скрыто присутствовать в категориях 1, 2, 4. С уверенностью идентифицировать текст, как сгенерированный мы можем, только если в сам текст включена информация о том, что он был создан конкретным ИИ, или если мы сгенерировали его сами. Во всех остальных случаях даже современные ИИ детекторы не могут со 100 % надежностью отличить тексты, созданные человеком, от текстов, созданных машиной. Это значит, что очень велика вероятность, что на занятиях мы будем иметь дело с текстами, сгенерированными ИИ, хотя не будем об этом знать или догадываться.

Одной из важных и необходимых характеристик, описывающих качество учебного текста является аутентичность, но имеется в виду не просто достоверность или подлинность текста. Проблема определения аутентичности и аутентичных материалов рассматривается в работах многих ученых, таких как Г. И. Воронина, А. А. Евтюгина, М. А. Казакова, К. С. Кричевская, Е. Ю. Лобанова, А. Мартинес, Р. П. Мильруд, Е. В. Носонович, А. А. Рогалева, Н. А. Тумакова [1; 2; 3; 4; 5; 6].

В Новом словаре русского языка Т. Ф. Ефремова определяет понятие «аутентичный» как «исходящий из первоисточника, соответствующий подлиннику; подлинный». При этом «подлинный» означает «являющийся оригиналом; не скопированный», а «оригинальный» представлен как «свойственный оригиналу, характерный для него» или «не заимствованный, не подражательный, не переводный, подлинный» [2, с. 51].

Многие методисты сходятся во мнении, что аутентичный текст – это документ, изначально не предназначенный для изучения иностранного языка. Рецепты, песни, опросы, прогнозы погоды, газетные статьи, рекламные странички, литературные тексты вошли в класс, чтобы служить примерами разнообразия письменных текстов, которые циркулируют в обществе [1, с. 137]. Г. И. Воронина определяет аутентичные тексты как тексты, заимствованные из коммуникативной практики носителей языка [5, с. 23–25].

Е. В. Носонович и Р. П. Мильруд рассматривают учебный текст, отвечающий нормам, принятым носителями языка, как текст, представляющий аутентичный дискурс, и выделяют следующие его характеристики:

1) культурологическая аутентичность – использование текстов, формирующих представления об основных особенностях жизни носителей изучаемого языка во всех сферах функционирования данного языка;

2) информативная аутентичность – использование текстов, содержащих актуальную, значимую информацию, соответствующую возрастным характеристикам обучающихся;

3) ситуативная аутентичность – создание естественной ситуации коммуникации, предложенной, например, в виде учебной иллюстрации, обладание заявленной темой интересом для носителей языка, естественность обсуждения указанной темой;

4) аутентичность национальной ментальности – разъяснение уместности или её отсутствия в использовании той или иной фразы иностранного языка;

5) реактивная аутентичность – способность текста вызывать у обучающихся аутентичный эмоциональный, мыслительный и речевой отклик;

6) аутентичность оформления – соответствие оформления текста в учебном пособии и его оформления в оригинале, имеющее следствием привлечение внимания учащихся, облегчение понимания коммуникативной задачи текста [6, с. 10–14].

Преимущества использования в обучении аутентичных текстов, по мнению А. А. Рогалевой, состоят в том, что это приводит к столкновению обучающихся «со всеми сложностями иностранного языка и культуры, они находятся в ситуации коммуникации, близкой к реальности и близкой к ситуации, которую переживает носитель языка, обращаясь к подобному документу в своей стране. Аутентичный текст как инструмент обучения иностранному языку мотивирует обучающегося, ставя его на место носителя языка и позволяя ему иметь прямой доступ к языку, как с точки зрения культуры, так и с точки зрения языка» [1, с. 137].

Недостатки аутентичных материалов и сложности в работе с ними описывает А. Мартинес. Особо отмечается сложность аутентичных текстов для понимания, связанная с различием культур, несоответствием уровня владения языком и используемых аутентичных языковых единиц, быстрое устаревание информации и, как результат, необходимость тщательного отбора материала и разработки процесса работы над ним [7].

Отдельно стоит упомянуть тот факт, что, как пишет А. А. Рогалева, тексты, созданные специально для учебных целей, воспроизводят многие элементы аутентичных текстов. Она соглашается с точкой зрения двух зарубежных лингвистов (D. Newby и J. P. Robert) в том, чтобы называть такую аутентичность поверхностной, а тексты – квазиаутентичными [3]. Кроме того, нельзя однозначно утверждать, что заимствованные из иноязычных первоисточников учебные языковые тексты не проходят никакой обработки, даже если текст не подвергался трансформации: попадая в дискурс урока, аутентичный текст изменяет свою прагматику [1, с. 137].

Можно ли считать сгенерированные ИИ иноязычные тексты аутентичными? Будут ли они обладать достоинствами аутентичных текстов, написанных людьми?

Как было сказано выше, в методике преподавания иностранных языков аутентичными принято считать тексты, заимствованные из коммуникативной практики носителей языка и изначально не предназначенные для изучения иностранного языка. Исходя из этого определения, тексты, сгенерированные ИИ не для обучения, а для реальных ситуаций общения в стране изучаемого языка, можно считать аутентичными. Если же текст был сгенерирован или трансформирован с помощью ИИ для образовательных целей, т.е. с учетом потребностей учащихся (тематики, объема, стиля, уровня владения языком, необходимости отработки определённых лексических или грамматических конструкций), он не является аутентичным. В обоих случаях не имеет принципиального значения, был ли текст сгенерирован по заданию носителя языка или преподавателя. Так, например, аутентичными будут считаться созданные ИИ кулинарный рецепт, описание маршрута, объявление о вакансии или краткое содержание статьи, или книги, если задание для генерации было составлено преподавателем иностранного языка как бы от лица жителя страны изучаемого языка, т.е. без поправок на цели обучения.

Таким образом, правильно использованный инструмент генеративных языковых моделей открывает обширное поле возможностей для актуализации учебных материалов, а также для дифференциации и индивидуализации обучения иностранным языкам. Но технологии искусственного интеллекта пока далеки от совершенства. Среди опасностей, связанных со сгенерированными текстами, наиболее значимой для обучения иностранным языкам является проблема «галлюцинаций» ИИ. Языковые модели могут генерировать неточные или ложные сведения («галлюцинации»), которые сложно выявить без глубоких экспертных знаний. Кроме того, нейросети не обладают здравым смыслом, что может приводить к генерации некорректных или вводящих в заблуждение текстов. Нам приходилось не раз сталкиваться с подобным явлением в процессе создания учебных текстов на иностранных языках. Действенным средством против этой проблемы является обязательная перепроверка фактов текста после его генерации. Это задание можно

поручить создавшему его ИИ или пропустить текст через другие, желательно, разные нейросети с целью проверить содержание на соответствие действительности.

ЛИТЕРАТУРА

1. Рогалева А. А. Концепция аутентичности учебного языкового текста" // Балтийский гуманитарный журнал. 2019. №1 (26). С. 136–138.
2. Казакова М. А., Евтюгина А. А. Аутентичные текстовые материалы в обучении иностранному языку // Вестник БГУ. Образование. Личность. Общество. 2016. №4. С. 50–59.
3. Рогалева А. А. Лингвистические и культурологические особенности учебных текстов (на материале немецкого и французского языков) // Научный журнал КубГАУ. 2015. №108. С. 1036–1046.
4. Лобанова Е. Ю., Тумакова Н. А. Аутентичность текста на занятии по иностранному языку // Молодой ученый. 2015. № 10 (90). С. 1201–1203. URL: <https://moluch.ru/archive/90/19101/>. (дата обращения: 26.06.2025).
5. Воронина Г. И. Организация работы с аутентичными текстами молодежной прессы в старших классах школ с углубленным изучением иностранного языка // Иностранные языки в школе. 1999. № 2. С. 23–25.
6. Носонович Е. В. Мильруд Г. П. Критерии содержательной аутентичности учебного текста // Иностранные языки в школе. 2008. № 2. С. 10–14.
7. Martinez A. Authentic Materials: An Overview // Karen's Linguistic Issues. 2002. URL: www.3.telus.net/linguisticsissues/authenticmaterials.html (date of access: 05.11.2003).

УДК 372.881.111.1'366:004.588

Листратова Ольга Константиновна

кандидат филологических наук,
доцент кафедры иноязычного
речевого общения
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Olga Listratova

PhD in Philology, Associate Professor
of the foreign communication
department
Belarusian State University
of Foreign Languages
Minsk, Belarus
atv-asp@mail.ru

ПРЫКЛАДНОЕ ПРАГРАМНАЕ ЗАБЕСПЯЧЭННЕ СПЕЦЫЯЛЬНАГА ПРЫЗНАЧЭННЯ ЯК СРОДАК АПТЫМІЗАЦЫІ ЗАСВАЕННЯ АНГЛАМОЎНАЙ ПОЛІМАРФЕМНАЙ ЛЕКСІКІ

Дадзены артыкул прысвечаны пытанням прымянення сучасных інфармацыйных тэхналогій у працэсе засваення англійскай мовы як замежнай. У артыкуле апісана распрацаванае аўтарам прыкладное праграмнае забеспячэнне спецыяльнага прызначэння, якое складаецца з комплексу заданняў, накіраваных на выкананне пяці тыпаў лінгвакагнітыўных дзеянняў па мадэляванню структуры полімарфемнай англамоўнай лексікі, яе семантызацыі, дыферэнцыяцыі, выяўленню парадыгматычных і сінтагматычных зносін англамоўных лексічных адзінак і іх марфалагічных элементаў. Праз апісанне вынікаў навучальнага эксперыменту раскрываецца патэнцыял распрацаванага праграмага забеспячэння для засваення полімарфемнай англамоўнай лексікі і паказваецца, як прымяненне камп'ютарных праграм дазваляе аптымізаваць працэс фарміравання замежнай маўленчай дзейнасці.

Ключавыя словы: прыкладное праграмнае забеспячэнне; камп'ютарны лінгватрэнажор; інтэрактыўнасць; англамоўная полімарфемная лексіка; семантызацыя; мадэляванне структуры лексічных адзінак; дыферэнцыяцыі лексічных адзінак.

APPLIED SOFTWARE FOR SPECIAL PURPOSES AS A MEANS OF OPTIMIZING THE ACQUISITION OF ENGLISH POLYMORPHEMIC VOCABULARY

The article is dedicated to the issues of applying modern information technologies in the process of mastering English as a foreign language. The article describes the applied software for special purposes developed by the author. The software consists of a set of tasks aimed at performing five types of linguo-cognitive actions related to modeling the structure of polymorphemic English words, their semantization, differentiation and identification of paradigmatic and syntagmatic relations established between English lexical units and their morphological elements. The description of the results of the undertaken experiment reveals the potential of the developed software for mastering polymorphemic English vocabulary and shows how the use of computer programs can optimize the process of developing foreign language speech activity.

Key words: applied software; computer linguistic trainer; interactivity; English polymorphemic vocabulary; semantization; modeling word structure; lexical units differentiation.

Падчас функцыянавання ў мове слова з'яўляецца не застылай, нязменнай адзінкай, а ўяўляе сабой адзінку рухомую, якой уласцівы шэраг фармальных і семантычных мадыфікацый. Аснову авалодання лексікай вывучаемай мовы складае засваенне сістэмы магчымых мадыфікацый. Базай разумення састаўленасці полімарфемнага слова з марфалагічных элементаў, што валодаюць значэннем і змяняюць такім чынам семантыку слова-асновы, з'яўляецца фарміраванне сістэмы лінгва-кагнітыўных дзеянняў, накіраваных на выяўленне ўзаемазалежнасці паміж семантычнымі і структурнымі характарыстыкамі слова. Вынікам засваення гэтай узаемасувязі індывідам выступае станаўленне ў яго здольнасці выконваць тры віды дзеянняў з полімарфемнымі словамі, менавіта 1) выводзіць семантыку полімафемнага слова, 2) мадэляваць структуру, якая адпавядае канкрэтнай сітуацыі маўлення, і 3) дыферэнцыраваць лексічныя адзінкі, якім характэрна агульнасць марфалагічнага элемента (кораню або афіксу).

Для забеспячэння асоб, якія вывучаюць англійскую мову як замежную, магчымасцю сфарміраваць здольнасці выконваць названыя вышэй лінгва-кагнітыўныя дзеянні намі было распрацавана даступнае для выкарыстання прыкладное праграмае забеспячэнне спецыяльнага прызначэння – камп'ютарны лінгватрэнажор ModeSt. Яго функцыянальнай асаблівасцю з'яўляецца накіраванасць на апрацоўку полімарфемных лексічных адзінак англійскай мовы праз раскрыццё значэнняў, якія закладзены ў іх марфалагічных элементах.

Выбіраючы спосаб рэалізацыі задуманага забеспячэння было вырашана, што ў дачыненні да пастаўленых задач неабходнасць вывучаць асяроддзі і пісаць праграму на мовах праграмавання адсутнічае. Такое прыкладное праграмае забеспячэнне спецыяльнага прызначэння можа быць рэалізавана з дапамогай інструментальных сродкаў агульнага праграмага забеспячэння Microsoft PowerPoint. Выкарыстанне яго тэхналагічных магчымасцей, такіх як а) гіперспасылкі, якія дазваляюць арганізаваць неабходны кантэнт у выглядзе гіпертэксту і адкрыць доступ да гіпертэксту менавіта ў тым месцы, якое карыстальнік лічыць патрэбным у кожны канкрэтны момант апрацоўкі слова,

б) трыггеры, якія ўяўляюць сабой аб'екты на экране, клік мышшу па якіх дае магчымасць ацаніць правільнасць сваіх дзеянняў [1, с. 87],

дазволіла распрацаваць дынамічны інтэрактыўны сродак засваення англамоўнай лексікі. Інтэрактыўнасць у дадзеным выпадку рэалізуецца праз прадстаўленне карыстальніку магчымасці самастойнага актыўнага дзеяння, калі, наводзячы мышшу на тыя ці іншыя элементы на экране, ён спрабуе, памыляецца, шукае і знаходзіць правільныя адказы, якія актывізуюць яго самастойную кагнітыўную дзейнасць і спрыяюць больш хуткаму і якасному засваенню новай лінгвістычнай інфармацыі.

У распрацаванае прыкладное праграмае забеспячэнне ўваходзіць комплекс з 29 серый заданняў пяці тыпаў, накіраваных на выкананне лінгва-кагнітыўных дзеянняў па:

- мадэляванню структуры англамоўнага слова з зададзенай семантыкай;
- семантызацыі полімарфемнага слова праз раскрыццё значэнняў яго марфалагічных элементаў;
- дыферэнцыяцыі лексічных адзінак, для якіх характэрны агульны марфалагічны элемент (корань або афікс);
- выяўленню парадыгматычных адносін, у якіх знаходзяцца лексічныя адзінкі;
- устанаўленню сінтагматычных адносін, у якія ўступаюць англамоўныя марфалагічныя элементы.

Заданні першага тыпу накіраваны на фарміраванне механізма мадэлявання структуры слова з неабходнай семантыкай і засваенне сінтагматычных сувязей марфем. На экране прадстаўлены простая невытворная англамоўная лексічная адзінка і яе пераклад на рускую мову. Справа прыведзены спіс з чатырох англамоўных афіксаў. Пры выбары карыстальнікам афікса адэкватнага для ўтварэння полімарфемнага слова яно з'яўляецца на экране. Пры няправільным выбары праграма паведамляе пра памылку.

Мэтай заданняў другога тыпу з'яўляецца навучанне сістэме лінгва-кагнітыўных аперацый, неабходных для вываду семантыкі вытворнага слова. Падчас выканання на экране прадстаўлены полімарфемная англамоўная лексічная адзінка і тры аднакарэнныя рускамоўныя лексічныя адзінкі, адна з якіх з'яўляецца карэктным перакладам першай. Для вызначэння слова-перакладу карыстальнік выконвае марфалагічны разбор англамоўнага слова і на падставе ведаў значэнняў марфем ажыццяўляе яго семантызацыю. Пры наведзенні мышшу на рускамоўныя лексічныя адзінкі праграма паведамляе пра карэктнасць выканання задання.

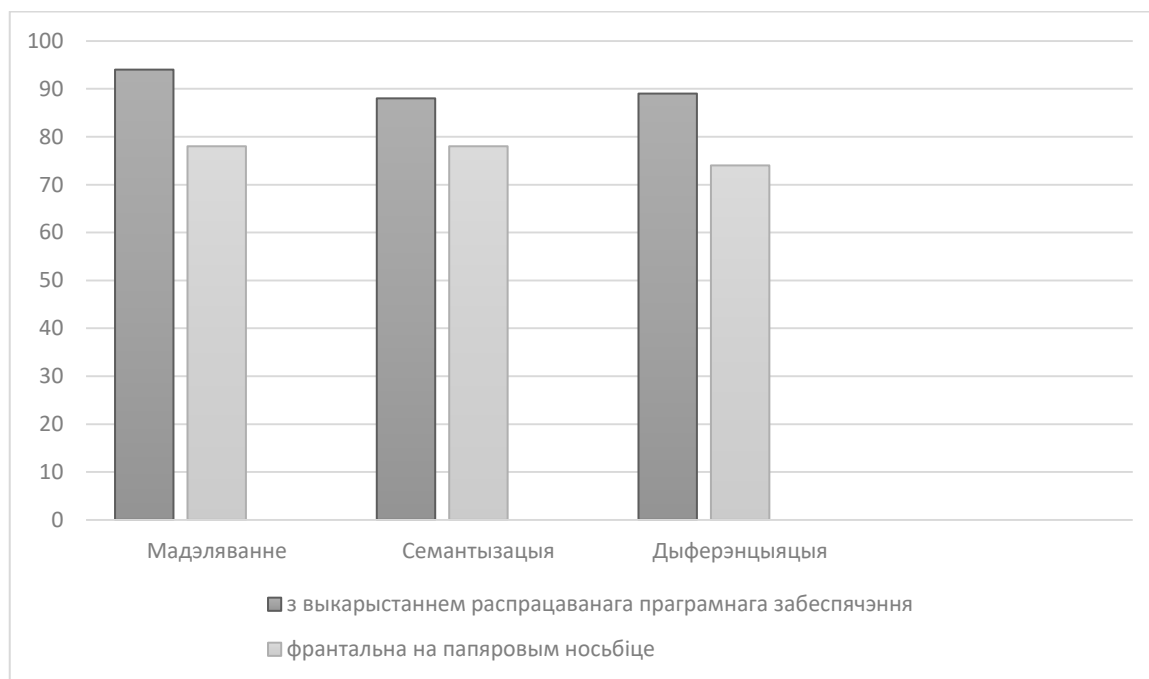
Выкананне заданняў трэцяга тыпу спрыяе засваенню карыстальнікам сінтагматычных адносін англамоўных марфем, канвенцый і асаблівасцей іх спалучальнасці. На экране прадстаўлены англамоўныя афікс і чатыры словы розных частак мовы. Выбар лексічнай адзінкі, якая можа спалучацца з прапанаваным афіксам, выклікае з'яўленне на экране полімарфемнага вытворнага слова. У выпадку выбару сінтагматычна несумяшчальнай з зададзеным афіксам каранёвай марфемы праграма паведамляе карыстальніку пра памылку.

Мэтамі заданняў чацвёртага тыпу выступаюць засваенне сінтагматычных адносін марфем і парадыгматычных адносін слоў, пабудаваных па розных словаўтваральных мадэлях, а таксама фарміраванне лінгва-кагнітыўных дзеянняў па мадэляванню структуры полімарфемнага слова. На экране зададзены невытворная англамоўная лексічная адзінка і спіс афіксаў (у адпаведнасці з лікам Мілера 7 ± 2). Пры кліку мышшу па афіксах, якія могуць выкарыстоўвацца для пабудовы вытворных ад зададзенага слова, на экране з'яўляюцца полімарфемныя словы і іх пераклад. Пры некарэктным выбары афікса праграма паведамляе карыстальніку пра памылку.

Заданні пятага тыпу накіраваны на пераадоленне інтэрферэнцыі роднай мовы праз фарміраванне англа-рускіх марфемных адпаведнасцяў. На экране прадстаўлены рускамоўны афікс, спіс англамоўных афіксаў, адзін з якіх адпавядае рускамоўнай марфеме, і англамоўнае невытворнае слова. Наводзячы мыш на англамоўныя афіксы, карыстальнік даведваецца пра правільнасць выканання задання.

Распрацаванае праграмнае забеспячэнне прайшло шасцігадовую апрабцыю і было ўкаранёна ў працэс навучання англійскай мове ў сярэдніх спецыяльных навучальных установах Рэспублікі Беларусь у 2019 годзе. У ходзе апрабцы для праверкі эфектыўнасці створанага праграмнага забеспячэння быў арганізаваны навучальны эксперымент, падчас якога была праведзена серыя ўрокаў з выкарыстаннем пяці інтэрактыўных серый лінгватрэнажора ModeSt.

Эксперымент паказаў, што фарміраванне лінгва-кагнітыўных дзеянняў па семантызацыі полімарфемных англамоўных лексічных адзінак, мадэляванні іх структуры і дыферэнцыяцыі лексічных адзінак, для якіх характэрны агульны марфалагічны элемент, з дапамогай распрацаванага комплексу адкрывае рэальныя перспектывы для авалодання полімарфемнай англамоўнай лексікай, паколькі дазваляе значна павысіць паказчыкі падыходных, дасягнутыя падчас працы фронтальна на папяровым носьбіце.



Параўнальны аналіз вынікаў падыходных па мадэляванні структуры, семантызацыі і дыферэнцыяцыі полімарфемных англамоўных лексічных адзінак, %

Як паказвае Рисунак, пры выкарыстанні распрацаванага праграмнага забеспячэння падыходныя дасягнулі паказчыка ў 94 % пры мадэляванні

структуры полімарфемнага англамоўнага слова, 88 % – пры яго семантызацыі і 89 % – пры дыферэнцыяцыі лексічных адзінак з агульным марфалагічным элементам. Атрыманыя паказчыкі значна перасягаюць паказчыкі, дасягнутыя пры працы фронтальна на папяровым носьбіце: 74–78 % ва ўсіх трох відах лінгва-кагнітыўнай дзейнасці. Атрыманне параўнальна высокага выніку пацвярджае эфектыўнасць прапанаванага праграмнага забеспячэння і сведчыць пра задавальненне з яго дапамогай патрэбаў навучэнцаў у лінгвістычна адаптаванай падачы ведаў.

Вынікі навучальнага эксперыменту паказалі, што адработка з дапамогай распрацаванага забеспячэння сістэмы лінгва-кагнітыўных дзеянняў па семантызацыі англамоўных лексічных адзінак, мадэляванню структуры полімарфемнага слова з зададзенай семантыкай, дыферэнцыяцыі лексічных адзінак, для якіх характэрны агульны марфалагічны элемент (корань або афікс), а таксама па выяўленню парадыгматычных і сінтагматычных адносін, у якіх знаходзяцца лексічныя адзінкі і іх марфалагічныя элементы, адкрывае рэальныя перспектывы для засваення лексікі англійскай мовы, паколькі:

1) дазваляе зрабіць працэс навучання інтэрактыўным і ўвесці ў яго элемент гульні, што рэалізуецца праз прадастаўленне навучэнцу магчымасці самастойнага актыўнага дзеяння;

2) садзейнічае аптымізацыі працэса фарміравання іншамоўнай маўленчай дзейнасці праз арганізацыю інтэрактыўнага дыялогу;

3) стварае магчымасць павысіць матывацыю навучання за кошт аперавання вывучаемымі аб'ектамі, выканання даступных заданняў з паралельнай візуалізацыяй вывучаемых лінгвістычных з'яў і аб'ектаў;

4) забяспечвае магчымасць самастойнага выбару рэжыму вучэбнай дзейнасці і даступнасць формаў прадстаўлення лінгвістычнага матэрыялу;

5) уключае кожнага ў працэс навучання і пашырае сферу самастойнай дзейнасці;

6) забяспечвае дыферэнцыраваны падыход і індывідуалізацыю навучання, якія ляжаць у аснове лінгвістычна-дыдактычных прыёмаў работы пры авалоданні замежнымі мовамі;

7) дазваляе пераадолець псіхалагічныя бар'еры, якія ўзнікаюць ў працэсе авалодання замежнымі мовамі, асабліва на пачатковых этапах;

8) фарміруе інфармацыйную культуру асабы праз прадастаўленне магчымасці паралельна засвойваць замежную мову і навучыцца выкарыстоўваць камп'ютар як рабочы інструмент.

ЛІТАРАТУРА

1. Листратова О. К. Структура и семантика англоязычной лексики в аспекте восприятия лицами с нарушением слуха и речи : дис. ... канд. филол. наук : 10.02.21. Минск, 2023. 178 л.

УДК 37.012:159.9.07

Мижевич Ольга Михайловна

кандидат философских наук, доцент,
доцент кафедры истории
и социальных наук
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Olga Mizhevich

PhD in Philosophy, Associate
Professor, Associate Professor
of the Department of History
and Social Sciences
Belarusian State University
of Foreign Languages
Minsk, Belarus
mizhevich.olga@mail.ru

ПСИХОЛОГИЧЕСКИЕ ПОДХОДЫ СТИМУЛИРОВАНИЯ УЧЕБНОЙ МОТИВАЦИИ СТУДЕНТОВ В ЦИФРОВОЙ ОБРАЗОВАТЕЛЬНОЙ СРЕДЕ

В данной статье проводится анализ ключевых психологических подходов, направленных на повышение учебной мотивации обучающихся, ставших особенно актуальными в условиях дистанционного и смешанного обучения и, которые получили широкое распространение в современном образовательном пространстве. Основываясь на теории самоопределения Эдварда Л. Деси и Ричарда М. Райана, теории атрибуции Бернарда Вайнера, а также мотивационной модели ARCS Харольда Келлера, выявляются психологические механизмы, способствующие удовлетворению базовых потребностей обучающихся: автономии, чувства компетентности и сопричастности. Подробно описываются практические методы стимулирования мотивации студентов, включая элементы геймификации, внедрение микролёрнинга и использование адаптивных цифровых учебных систем. Также предлагаются рекомендации по повышению эффективности курсов иностранного языка с учётом потенциала современных цифровых технологий. В статье оценивается роль методик саморегуляции, техник постановки целей, а также возможностям мониторинга и оценки личностного прогресса, которые играют решающую роль в поддержании устойчивой учебной мотивации и формировании у студентов активной учебной позиции.

Ключевые слова: мотивация; цифровое обучение; теория самоопределения; теория атрибуции; модель ARCS; геймификация; микролёрнинг; адаптивность.

PSYCHOLOGICAL APPROACHES TO STIMULATING STUDENTS' LEARNING MOTIVATION IN A DIGITAL EDUCATIONAL ENVIRONMENT

This article analyzes key psychological approaches aimed at increasing students' motivation for learning, which have become especially relevant in the context of distance and blended learning and have become widespread in the modern educational space. Based on Edward L. Deci and Richard M. Ryan's self-determination theory, Bernard Weiner's attribution theory, and Harold Keller's ARCS motivational model, the psychological mechanisms that help meet students' basic needs are identified: autonomy, a sense of competence, and belonging. Practical methods for stimulating students' motivation are described in detail, including elements of gamification, the introduction of microlearning, and the use of adaptive digital learning systems. Recommendations are also offered for increasing the effectiveness of foreign language courses, taking into account the potential of modern digital technologies. The article assesses the

role of self-regulation methods, goal setting techniques, and the possibilities of monitoring and evaluating personal progress, which play a decisive role in maintaining sustainable learning motivation and developing an active learning position in students.

Key words: motivation; digital learning; self-determination theory; attribution theory; ARCS model; gamification; microlearning; adaptability.

Стремительное развитие цифровых технологий и глобальная пандемия значительно увеличили объем дистанционного и смешанного формата в образовательном процессе. При традиционной очной (дневной или вечерней) модели преподаватель мог «вживую» регулировать темп занятия, поддерживать внимание слушателей и оперативно реагировать на их реакцию. Что касается цифровой образовательной среды, то в данном случае студенты стали чаще сталкиваться с дефицитом внешних стимулов, сложностями самоконтроля и уменьшением чувства принадлежности к учебному сообществу. В настоящее время это обостряет задачу поиска обоснованных педагогических и психологических методов, направленных на удержание интереса и повышение ответственности учащихся за собственное обучение.

Цель данной работы – проанализировать научные психологические подходы формирования учебной мотивации и практические методики создания обучающих курсов, способствующие созданию благоприятной цифровой образовательной среды. В настоящее время профессорско-преподавательский состав языковых высших учебных заведений нуждается в научно обоснованных методах, позволяющих эффективно вовлекать студентов в учебный процесс, повышать их активность и ответственность за результаты. В статье систематизируются психологические подходы к мотивации в онлайн и предлагает практические алгоритмы для дизайна курсов иностранного языка. Акцентируется внимание на особенностях лингвистического образования, что позволит преподавателям языковых вузов адаптировать предложенные подходы и методики к конкретным учебным ситуациям и технологическим платформам.

В современной психологии мотивации наиболее ценный теоретический и практический инструментарий предлагается в теории самоопределения, разработанной Эдвардом Деси и Ричардом Райаном (1985). Данную теорию хорошо дополнили и детализировали еще две: теория атрибуции Бернарда Вайнера (1985) и в мотивационной модели ARCS Гарольда Келлера (1987).

В рамках статьи целесообразно более подробно остановимся на рассмотрении теории самоопределения (Self-Determination Theory, SDT). Это объясняется тем фактом, что в настоящее время данная теория является основным методологическим ориентиром в исследовании личностного и профессионального развития, предлагая целостный подход к мотивации как динамической системе, зависящей от качества взаимодействия между индивидом и обществом. В образовательном процессе SDT активно используется для выявления факторов, способствующих формированию устойчивой учебной мотивации, что особенно актуально в условиях цифровой педагоги-

ческой среды. Э. Деси и Р. Райан утверждают, что внутренняя мотивация усиливается при удовлетворении трёх базовых психологических потребностей:

- автономия, предполагающая наличие у человека ощущения свободы выбора действий и способов их выполнения;
- компетентность, говорящая об уверенности в своих возможностях и ощущение прогресса;
- сопричастность, которая означает наличие значимых социальных связей с преподавателем и сверстниками [1].

Актуальной при исследовании мотивационных процессов остается *теория атрибуции*, сформулированная Фритцем Хайдером. Её применение позволяет глубже понять, как интерпретации причин поведения влияют на эмоции, решения и развитие личности. Особое значение для повышения эффективности образовательного процесса имеет адаптивная модель Б. Вайнера. В его модели причины поведения индивида классифицируются по трем параметрам: *локус контроля* – внутренняя или внешняя причина; *стабильность* – постоянство причины во времени; *контролируемость* – возможность воздействия на причину. Эти параметры влияют на эмоциональные реакции, например, чувство гордости, вины или стыда, и определяют поведенческие стратегии, такие как повторение действия или избегание.

Б. Вайнер в своей концепции атрибутивной теории акцентирует внимание на восприятии студентами причин успехов и неудач, которые оказывают влияние на уровень мотивации и самооценку [2]. Так, приписывание успеха внутренним, стабильным и контролируемым факторам способствует формированию высокой мотивации. И наоборот, приписывание неудач внутренним, неконтролируемым причинам, например, «низким способностям», может вызвать амотивацию и тревожность. Перенос атрибуции в зону контролируемых факторов (усилия, навыки) способствует развитию настойчивости и снижению тревожности перед экзаменами.

В своей *мотивационной модели ARCS*, которая наиболее пригодна для проектирования языковых курсов, Г. Келлер выделил четыре мотивационных компонента, которые должны быть представлены в образовательном процессе:

attention – привлечение и удержание внимания обучающихся через элементы новизны, разнообразие и неожиданность (предполагает применение визуальных и аудиальных стимулов);

relevance – демонстрация значимости материала для личных целей и потребностей обучающихся (использование примеров из культуры и жизни стран изучаемого языка);

confidence – формирование у обучающихся ощущения контроля над обучением и веры в собственный успех (реализуется благодаря пошаговому усложнению задач и обратной связи);

satisfaction – предоставление позитивного подкрепления и условий для ощущения достижения (создание эмоционально насыщенной среды обучения, применение геймификации).

Изучение иностранного языка требует высокой степени вовлечённости, регулярной практики, самоконтроля и преодоления коммуникативных барьеров. При этом мотивация обучающихся часто подвержена колебаниям, особенно в условиях дистанционного и смешанного обучения. Все три мотивационные теории могут выступать как основа для разработки языковых программ, способных поддерживать интерес и стимулировать обучение.

На практике в образовательном процессе были разработаны целый ряд механизмов, которые за относительно короткий срок доказали свою эффективность. Рассмотрим наиболее известные и действенные в языковой образовательной среде.

Для выполнения заданий достаточно давно используют письменные эссе, аудиоподкасты, видеопрезентации. Ведение языковых дневников, анализ ошибок, постановка личных целей призваны формировать навыки рефлексии, формирует у студентов ощущение контроля над образовательным процессом и повышают степень их вовлечённости. Хорошо зарекомендовал себя модульный дизайн причем без использования жёсткой навигационной иерархии, позволяющий студентам перемещаться между темами по собственному усмотрению, а также гибкая система дедлайнов, адаптирующаяся к индивидуальному ритму работы. Необходимо уточнить. Что при использовании таких механизмов преподаватель выступает не как носитель знания, а как посредник и консультант, создающий условия для самостоятельного обучения.

Внедрение микролёрнинга, который проявил себя в виде кратких тестовых заданий с мгновенной автоматической обратной связью и подробным разбором ошибок, значительно снизил когнитивную нагрузку на студента, что способствует лучшему усвоению учебного материала. В микролёрнинге чаще всего используются такие цифровые платформы как EdApp, Duolingo, Quizlet, Coursera Shorts.

В условиях цифрового обучения особую роль играет проектирование заданий, обеспечивающих обучающимся возможность выбора, самоконтроля и рефлексии. Одним из эффективных решений стало внедрение *адаптивных модульных систем*, позволяющих студентам самостоятельно определять уровень сложности и маршрут прохождения учебного материала [3]. Такие модульные системы обеспечивают гибкость, поддерживают чувство автономии и способствуют формированию индивидуальных траекторий обучения. Например, использование *ветвящихся сценариев* в цифровых курсах позволяет студенту выбирать последовательность изучения тем, получать обратную связь и корректировать собственную стратегию освоения материала. Это не только усиливает мотивацию, но и развивает навыки саморегуляции, критического мышления и ответственности за результат.

Укрепление у студентов потребности сопричастности способствует организации групповых проектов и парных заданий в виртуальных комнатах (Zoom, Microsoft Teams) с возможностью совместного редактирования; активное использование форумов, чатов и социальных платформ для обмена

идеями, совместного обсуждения ошибок и взаимопомощи; регулярные онлайн-консультации и «виртуальные кабинеты» преподавателя для поддержки и обратной связи [4]. Данные практики способствуют формированию учебного сообщества и предотвращают чувство изоляции.

В работе со студентами важно формировать позитивные атрибуции, способствующие саморазвитию и самооффективности. Применение атрибутивных рефреймингов может повысить уровень академической вовлеченности и снизить проявления выученной беспомощности. В цифровой среде автоматические тесты с объяснением ошибок и предложениями альтернативных стратегий обучения помогают студентам связывать результаты с собственными усилиями, а не с «внешними» или «неизменными» факторами [5]. Это повышает устойчивость к неудачам и стимулирует более активное обучение.

Примером интеграции модели ARCS в языковое обучение может служить создание языковых квестов, в которых студенты проходят задания, построенные по принципам ARCS, получают баллы, обратную связь и возможность выбрать формат обучения. Перспективным направлением является адаптация ARCS-модели к онлайн-платформам и обучающим мобильным приложениям, а также интеграция с технологиями искусственного интеллекта, что позволит создать индивидуальные схемы овладения языком.

Современное образование, и лингвистическое в том числе, все чаще обращается к идее использования механизмов геймификации. Геймифицированные подходы позволяют создать адаптивную среду, повышающую интерес к языковому материалу и облегчая процессы усвоения и повторения.

Геймификация основывается на теориях мотивации Э. Деси и Р. Райана, Г. Келлера, когнитивного обучения и теории активного участия. Примерами использования геймификации в образовательном процессе могут послужить цифровые платформы и приложения (Duolingo, LingQ, Wordwall), предлагающие визуализированные задания с элементами награды в виде дигитальных бейджей и сертификатов за прохождение ключевых модулей и достижение языковых уровней, а также лексические квесты и морфологические лабиринты, которые стимулируют студентов к повторению материала через поиск терминов, грамматических форм. Хорошо зарекомендовали себя создание лидербордов и проведение командных соревнований на основе статистики грамматических викторин; моделирование дискурса и игровых ситуаций, где обучающиеся разыгрывают коммуникативные роли: переводчики, эксперты, медиаторы.

Геймификация представляет собой достаточно перспективный подход в лингвистическом образовании, способный трансформировать традиционные методики обучения и сделать языковое развитие более увлекательным, осознанным и эффективным. При правильном педагогическом сопровождении игровые элементы способствуют глубокому овладению лингвистическими концептами и развитию критического мышления студентов. Но, несмотря на

преимущества, при этом важно учитывать возможность появления у обучающихся игровой переутомляемости и поверхностного освоения материала, а также сложности оценивания и переноса геймифицированных результатов в десятибалльную академическую систему.

Интегрированное применение описанных методов позволяет не только повысить краткосрочную мотивацию, но и выработать устойчивые учебные привычки. Ведут к формированию у студентов устойчивой внутренней мотивации, повышению когнитивной активности, уменьшению уровня тревожности при говорении на иностранном языке, повышению академической успешности и самостоятельности студентов. Важно соблюдать баланс между внешними стимулами (геймификация, лидерборды) и внутренними стимулами (самооценка, смысловая значимость контента). Чрезмерная зависимость от балльной системы может способствовать поверхностному освоению материала, тогда как акцент на самоопределении и саморегуляции поддерживает глубокую мотивацию и профессиональный рост будущих преподавателей иностранных языков.

Современная цифровая трансформация образования требует от преподавателей овладения передовыми психологическими методами стимулирования мотивации. Комбинация поддержки автономии, развития компетентности и укрепления сопричастности, а также использование микролёрнинга и геймификации, создаёт благоприятные условия для высокой вовлечённости студентов. В результате они не только совершенствуют языковые навыки, но и приобретают ключевые самоорганизационные компетенции, необходимые для успешной педагогической деятельности в новых образовательных реалиях.

ЛИТЕРАТУРА

1. Деси Э., Райан Р. Теория самоопределения и мотивация учащихся. М. : Академический проект, 2020. 290 с.
2. Вайнер Б. Мотивация и эмоция. СПб.: Питер, 2018. 800 с.
3. Свиридова О. В. Цифровая педагогическая компетентность преподавателя: вызовы и перспективы. // Высшее образование в России. 2022. № 4. С.10–14
4. Радчикова Н. П., Одинцова М. А., Сорокова М. Г. Психологические факторы отношения студентов к цифровой образовательной среде (на примере российских и белорусских вузов) URL: <https://cyberleninka.ru/article/n/psihologicheskie-factory-otnosheniya-studentov-k-tsifrovoy-obrazovatelnoy-srede-na-primere-rossiyskih-i-belorusskih-vuzov> (дата обращения: 01.07.2025).
5. Ярычев М. У. Формирование устойчивых мотивационных установок у студентов в условиях цифровой перегрузки URL: <https://cyberleninka.ru/article/n/formirovanie-ustoychivyh-motivatsionnyh-ustanovok-u-studentov-v-usloviyah-tsifrovoy-peregruzki>. (дата обращения: 01.07.2025).

УДК 800

Петрова Марина Георгиевна
кандидат педагогических наук,
доцент
Российский университет дружбы
народов имени Патриса Лумумбы
г. Москва, Россия

Petrova Marina Georgievna
PhD in Pedagogy, Associate Professor
Peoples' Friendship University
of Russia (RUDN University)
Moscow, Russia
petrova-mg@rudn.ru

НЕЙРОКОГНИТИВНЫЕ ОСНОВЫ АДАПТИВНОГО ОБУЧЕНИЯ С ИИ

Современная образовательная парадигма претерпевает значительные изменения под влиянием развития искусственного интеллекта (ИИ) и новых открытий в области нейрокогнитивных механизмов обучения. В данной статье рассматриваются ключевые зоны мозга, вовлеченные в процесс овладения иностранными языками, и их роль в адаптивном обучении. На основе нейробиологических данных предложены педагогические стратегии, оптимизирующие усвоение языков с учетом когнитивных факторов, таких как рабочая память, когнитивный контроль и метапознание. Особое внимание уделено инновационным образовательным проектам Института иностранных языков РУДН, включающим: адаптивную платформу Edutainment Chinese (EC), сочетающую цифровые технологии, культурное погружение и ИИ для изучения китайского языка; виртуальные тренажеры «VR-тур в Древний Китай» и «Global Negotiation Quest», направленные на создание иммерсивной среды для освоения китайского и английского языков. В статье также предложены перспективные направления разработки: нейрокогнитивно совместимых ИИ-систем, интегрирующих современные лингводидактические подходы, нейромониторинг и искусственный интеллект; «когнитивных цифровых двойников» – персональных ИИ-ассистентов, адаптирующихся к индивидуальным особенностям учащихся. Результаты исследования демонстрируют потенциал нейронауки и ИИ в трансформации языкового образования, обеспечивая персонализацию обучения и повышение его эффективности.

Ключевые слова: адаптивное обучение иностранному языку; искусственный интеллект в образовании; нейрокогнитивные механизмы обучения; персонализированное языковое обучение; нейробиология изучения второго языка.

NEUROCOGNITIVE FOUNDATIONS OF ADAPTIVE FOREIGN LANGUAGE TEACHING WITH AI

The modern educational paradigm is undergoing significant changes due to the development of artificial intelligence (AI) and new discoveries in the field of neurocognitive learning mechanisms. This article examines the key areas of the brain involved in the process of learning foreign languages and their role in adaptive learning. Based on neurobiological data, pedagogical strategies are proposed to optimize language acquisition, taking into account cognitive factors such as working memory, cognitive control, and metacognition. Special attention is given to innovative educational projects at the RUDN University Institute of Foreign Languages, including the adaptive Edutainment Chinese (EC) platform, which combines digital technologies, cultural immersion, and AI for learning Chinese; as well as virtual simulators like "VR Tour to Ancient China" and "Global Negotiation Quest," aimed at creating an immersive environment for learning Chinese and English. The article also suggests promising areas for

development: neurocognitively compatible AI systems that integrate modern linguistic and didactic approaches, neuromonitoring, and artificial intelligence; and "cognitive digital twins" – personal AI assistants that adapt to the individual characteristics of students. The results of this study demonstrate the potential of neuroscience and AI in transforming language education by providing personalized learning experiences and improving overall effectiveness.

Key words: foreign language acquisition; artificial intelligence in education neurocognitive learning mechanisms; personalized language instruction; neurobiology of second language learning.

The relevance of this study stems from the necessity to transition from traditional to personalized education models in the digital era, the rapid advancement of AI technologies that often overlook students' neurocognitive profiles, the demand for evidence-based methods to enhance learning efficiency, and the widening gap between technological capabilities and their pedagogical implementation.

Neuro-linguistics, which examines the brain mechanisms underlying language acquisition, can significantly benefit from AI-driven data analysis and computational modeling to deepen our understanding of linguistic processes. AI enables the adaptation of educational content to individual learners' cognitive profiles, aligning with neuro-linguistic principles that acknowledge variability in language learning across individuals [1, p. 51]. Furthermore, AI-enhanced language instruction can substantially improve learning outcomes by providing students with previously inaccessible tools for practice and real-time feedback.

Contemporary neuro-linguistic didactics increasingly intersects with psycholinguistics and cognitive science, particularly within the framework of the St. Petersburg School led by Tatiana Chernigovskaya [2]. Neuro-linguistics operates at the intersection of four disciplines:

- Neuropsychology & Neurobiology – investigating brain structures and functions;
- Linguistics – analyzing how the brain processes, produces, and structures language;
- Didactics – a branch of pedagogy focused on learning methodologies.

Numerous neuro-linguistic studies carry out a comparative analysis of the structural features of the brain in different groups of participants. In particular, the differences between monolinguals and bilinguals are studied [3,4], and changes in the brain of second language learners (L2) are monitored as their language skills improve [5; 6; 7].

Factors such as professional linguistic experience (e.g., in translators or language teachers), residence in multilingual regions with two or three official languages, and engagement in cognitively stimulating activities have been shown to induce neuroplastic changes. These include structural modifications in the corpus callosum and variations in gray and white matter density and connectivity in specific cortical regions of both cerebral hemispheres [8; 9; 10].

Voxel-based morphometry (VBM) occupies a special place among neuroimaging methods, which makes it possible to compare gray matter density between different groups and over time in the same subjects based on structural MRI (magnetic resonance imaging) data [6].

In a review by Green et al. [11] presents the following key findings:

In English-speaking monolinguals, a positive correlation was found between the volume of gray matter in the posterior supra marginal gyrus (SMG) and the size of vocabulary.

Italian-English bilinguals showed greater gray matter density in the same area (posterior SMG) compared to monolinguals. In Chinese speakers, in addition to an increase in density in the posterior SMG, structural changes were observed in:

- the right superior temporal gyrus;
- the right inferior frontal gyrus (RFG);
- the left middle and upper temporal gyri.

These findings suggest that:

Language experience causes specific structural changes in the brain. Firstly, the localization of changes may vary depending on the linguistic characteristics of the native language. Secondly, the posterior marginal gyrus plays a special role in lexical processing, regardless of language status.

Methodological clarifications:

VBM analyzes the concentration of gray matter at the voxel level (3D pixels). Studies include both cross-sectional (group comparison) and longitudinal (change tracking) designs.

The data obtained are consistent with the hypothesis of neuroplasticity induced by language learning.

Particular attention should be given to neural plasticity in language acquisition, where “cortical regions demonstrate differential activation patterns when processing a new language compared to the native language. Linguistic features such as orthographic systems and tonal/atonal characteristics influence the neurocognitive foundations of language learning through neural specialization” [12].

As evidenced by contemporary research, foreign language acquisition constitutes a complex cognitive process requiring integrated activation of multiple neural networks. Neurobiological studies provide crucial insights into: (1) the cerebral mechanisms underlying new language learning, (2) the specific brain regions activated during various linguistic tasks, and (3) evidence-based approaches for optimizing the learning process [13].

The principal neuroanatomical structures involved in foreign language processing are systematically organized. As presented in Table 1, linguistic information follows a defined neural pathway from perceptual processing (Wernicke's area) to speech production (Broca's area), with grammatical integration mediated by the arcuate fasciculus.

Table 1

Neural Structures and Their Role
in Foreign Language Teaching

Neural Structure	Localization	Function in FLT
Broca's area	Inferior frontal gyrus	Speech production, grammar, syntax
Wernicke's area	Superior temporal gyrus	Understanding the meaning of words and speech
arcuate fasciculus	Connects Broca's and Wernicke's zones	Grammar mastering
Right hemisphere	The right hemisphere of the cortex	Intonation, prosody, understanding of context
basal ganglia	Subcortical structures	Automation of speech skills (fluency)
Hippocampus	Temporal lobes	Memorizing new vocabulary (episodic memory)
prefrontal cortex	Frontal lobes	Attention control, switching from one language to another
Amygdala	in the white matter of the temporal lobe of each hemisphere (1.5–2.0 cm behind the temporal pole)	Stress response

Language acquisition involves not solely cortical processing, but also engages subcortical structures (including the basal ganglia and hippocampus) and regulatory systems (particularly the prefrontal cortex).

Neuroplasticity research demonstrates that grammatical proficiency development requires strengthening white matter connectivity between Broca's area and Wernicke's area via the arcuate fasciculus. This neuroanatomical evidence supports the pedagogical effectiveness of simultaneous auditory processing and speech production exercises.

Fluency acquisition in second language learning relies significantly on basal ganglia circuitry, which facilitates procedural memory formation. Evidence-based approaches include:

- Shadowing techniques (real-time vocal imitation of native speech);
- Chunking methods (automatized phrase learning);
- These strategies reduce cognitive load through neural automation.

Spaced repetition systems (e.g., Anki) promote speech automation by optimizing memory consolidation through the hippocampus and establishing procedural memory pathways in the basal ganglia. This dual-system engagement enhances long-term retention and fluency.

Effective listening is achieved by activating the Wernicke's area through the analysis of the meaning of words (for example, by listening to adapted podcasts that include vocabulary analysis) and by engaging the right hemisphere to recognize the emotional nuances of speech (such as listening to dialogues with varying intonations). Thus, using audio materials with emotionally charged speech (like podcasts and interviews) develops not only vocabulary (associated with the Wernicke's area) but also the ability to grasp hidden meanings (linked to the right hemisphere).

The hippocampus, which is involved in episodic memory, plays a crucial role in memorizing new vocabulary. However, for words to be retained long-term, it is essential to connect vocabulary with personal experiences (for example, by creating associations with vivid images or events) and to employ spaced repetition methods. This is because the hippocampus encodes information more effectively with gradual consolidation. While learning a foreign language, it is advisable to use applications that incorporate interval repetition (such as Memrise), which are effective for long-term retention, as they align with the functioning of the hippocampus.

Broca's area is responsible for mastering grammar and syntax, while the arcuate fasciculus facilitates the connection between comprehension and production. Grammatical rules are learned more effectively when theory is combined with practice (for instance, by immediately applying rules in spoken language) or through visualization, as sentence patterns activate Broca's area. A foreign language teacher should recognize that grammar exercises should include not only gap-filling activities but also the construction of oral utterances. This approach strengthens the arcuate fasciculus and accelerates the application of rules in speech.

The amygdala in each hemisphere and the prefrontal cortex, which regulates attention, are responsible for reducing stress when speaking. Stress can inhibit or disrupt the cognitive functioning of both children and adults. Therefore, it is vital to create a safe environment that promotes a pleasant release of endorphins. This involves fostering a positive attitude, developing a constructive self-perception, and ensuring that teachers recognize the uniqueness of each student, including their cognitive styles and intellectual potential. It is also essential to facilitate collaboration and group work, both of which enhance awareness and reduce tension [14, p. 65].

The teacher needs to remember that the fear of making mistakes can block language development, so it is important to:

1. Train improvisation (for example, through games like “5 Minutes of Talking Without Stopping”).

2. Reduce anxiety through positive reinforcement (praising effort, not just results).

For example, role-playing in a safe environment (such as the classroom) reduces the activation of the amygdala and enhances the control of the prefrontal cortex over speech processes.

The interpretation outlined above allows us to move from theory to specific pedagogical strategies based on neuroscience. Table 2 demonstrates the relationship between neuroscience and the practice of language education.

Table 2
Learning Strategies Based on Neuroscience

Learning objective	Key Neuro Structure	Recommended methods
Listening	Wernicke’s area + right hemisphere	emotional speech listening
Grammar	Broca’s area + arcuate fasciculus	visualization + oral practice
Vocabulary memorizing	Hippocampus	interval repetitions + associations
Verbal fluency	Basal ganglia	shadowing, memorizing speech patterns
Stress reduction	Prefrontal cortex	role-playing games, positive reinforcement

This study reveals that contemporary foreign language educators must develop a multidimensional competency profile encompassing:

Core Teaching Competencies:

- Advanced proficiency in target language communication.
- Theoretical mastery of language teaching methodologies.

Neurobiological Knowledge:

- Understanding of cognitive load theory.
- Working memory capacity limitations.
- Neural mechanisms of language acquisition.

Technological Integration:

- Collaboration with AI systems to develop adaptive learning platforms.
- Implementation of personalized learning trajectories.
- Digital tools for cognitive load optimization (Table 3).

Empirical evidence demonstrates the pedagogical superiority of brief, interactive instructional sessions over extended lectures for learners with limited attention capacity (Sweller, 2020). This cognitive-adaptive approach enables educators to:

- Align teaching methods with neurobiological principles of language learning.
- Accommodate individual neurocognitive profiles.
- Optimize knowledge acquisition through evidence-based practices.

Table 3

The influence of cognitive factors on aspects of language

Cognitive factor	Lexic	Грамматика	Произношение	Аудирование
Working memory	Запоминание новых слов	Усвоение сложных структур	Имитация звуков	Удержание услышанного
Cognitive control	Подбор синонимов	Применение правил	Коррекция акцента	Фильтрация шумов
Метапознание	Выбор стратегий запоминания	Анализ ошибок	Самооценка артикуляции	Выделение ключевых слов

The Institute of Foreign Languages at RUDN University pays great attention to the development of digital educational platforms, virtual simulators, smart manuals, and AI-based applications tailored to the cognitive processes of students. All digital content is created in close cooperation with the student scientific circle “Digital Professor” and graduate students of the Institute.

The modern educational platform Edutainment Chinese (EC) offers a comprehensive digital approach to learning Chinese, combining advanced technology with cultural immersion. A key feature of EC is the use of immersive VR technologies that allow students to take virtual tours of the famous Gugun Museum. This creates a sense of real presence and cultural immersion, significantly increasing motivation to learn. The platform integrates the latest generative AI models, including ChatGPT and deep learning algorithms, which provide personalized conversational practice, adaptive learning, and instant feedback. Interactive Quizlet exercises, an animated hieroglyphic writing learning system, and gamified learning elements are utilized to effectively memorize hieroglyphs and grammar.

Special attention is given to authentic materials, including a selection of Chinese films with subtitles, popular musical compositions, and literary works. The platform includes an intelligent speech recognition system that helps improve pronunciation and offers automated preparation for the HSK international exam with adaptive tests.

EC provides full integration with popular Chinese digital resources: The Baidu search engine, the NetEase Cloud Music platform, and the Chinese Plus and Chaoxing Learning System educational portals.

This comprehensive approach makes Edutainment Chinese a unique solution for effective and engaging learning of the Chinese language and culture.

Post-graduate students of the IFL RUDN University participate in the creation of virtual simulators that incorporate elements of artificial intelligence. The virtual simulator “VR Tour to Ancient China” was developed for a seminar session conducted using virtual reality in the discipline “Methods of Teaching Foreign Languages and Cultures.” The purpose of this product is to introduce new pedagogical technologies in the era of digitalization for teaching foreign languages and cultures through a virtual journey to Ancient China. In developing this virtual product, graduate students created visual content using neural networks such as Midjourney, Kandinsky, and Stable Diffusion. Some ideas for creating locations in virtual space were generated by the ChatGPT-4 models. The latest location, “The Chinglish Test,” was created using a 360-degree aerial view of Shanghai. Dialogues in Chinese are recorded by native speakers.

The next project, Global Negotiation Quest VR simulator, focuses on creating an innovative educational environment that helps students immerse themselves in the study of socio-political contexts in various regions using virtual reality (VR) technologies. Its goal is to enhance language competence in English and to develop a systematic understanding of the interrelationships between language, culture, and social structures, thereby contributing to the formation of intercultural communication and analytical skills.

The theme “Arrivals. Visits. Talks” is explored as part of the introductory block of the discipline “Language of the Region of Specialization” and aims to familiarize students with the protocol and etiquette of official visits. In the process of completing seminar tasks, students acquire foundational knowledge and skills in verbal and non-verbal communication, establishing diplomatic relations, and conducting international negotiations. They immerse themselves, with the help of neural networks and end-to-end digital technologies, into a completely new educational space for teaching a foreign language – essentially their own metaverse for the language being studied [15, p. 64].

In the future, the development of neurocognitively compatible AI systems represents a qualitatively new class of educational technologies that take into account the principles of the human brain when learning languages. These systems are based on the deep integration of three key components: next-generation artificial intelligence, neurocognitive monitoring data, and modern linguistic and didactic approaches. While such systems already exist (e.g., <https://neuralink.com/>), they have not yet been applied in the field of learning foreign languages.

Imagine the amazing results that could be achieved if students had neural gadgets for tracking their cognitive processes. These gadgets could monitor the activation of Broca's area during grammatical exercises, the active work of the hippocampus in the process of memorizing vocabulary, and the involvement of the right hemisphere in learning intonation.

Based on this data, the algorithm can dynamically adjust:

- The difficulty and type of exercises;

- The rate of material delivery;
- The ratio of different types of speech activity;
- The emotional tone of the educational content.

When analyzing stress indicators, a gradual decrease in anxiety associated with speaking a foreign language can be observed. This can be facilitated through virtual dialogues with gradually increasing complexity, real-time biofeedback, and neurostimulating techniques based on the principles of cognitive behavioral therapy.

Another promising area is the development of “cognitive digital twins” – personal AI assistants that accompany students for years, studying their cognitive styles, language acquisition patterns, individual “blind spots” in language learning, and optimal memorization strategies.

Such systems will completely transform the language education landscape in the next 5–10 years, enabling truly personalized learning that takes into account not only the level of language proficiency but also the unique neurocognitive characteristics of each student. This will not only accelerate the process of language acquisition by 2–3 times but also make it more natural, aligning with how the human brain is evolutionarily adapted to learn new languages.

LITERATURE

1. Петрова М. Г. Перспективы взаимодействия нейролингвистики и обучения иностранным языкам в условиях развития ИИ. Международные Ахановские чтения – XXVII «ТЮРКСКИЙ МИР: ЯЗЫК, ИСТОРИЯ И КУЛЬТУРА». 31 мая-01 июня 2024 г. – Алматы: Казахский национальный университет им. Аль-Фараби, 2024. – С. 51–56.
2. Черниговская Т. В. Чеширская улыбка кота Шредингера: язык и сознание. М.: Языки славянской культуры, 2013. 448 с.
3. Шишковская Т. И. Речевой коннектом // Психиатрия и нейронауки [Electronic resource] 2020. URL: <https://psyandneuro.ru/kursy/nejrolingvistika/contemporary-model-of-language-organization/> (date of access: 10.07.2025).
4. Шишковская Т. И. Нейроанатомия речи: от Брока и Вернике к фонологии и семантике // Психиатрия и нейронауки [Electronic resource]. 2019 URL: <https://psyandneuro.ru/kursy/nejrolingvistika/neuroanatomy-of-language/#more-3732> (date of access: 10.07.2025).
5. Пашкевич С. Нейролингвистика на стыке наук // Наука инновации. 2023. №7 (245). С. 6–12.
6. Pandža Nick. Collecting and Analyzing L2 Data on Neurocognitive Mechanisms // Eliciting Second Language Data. 2023. Pp. 233–254. 10.1002/9781394259670.ch11.
7. Zhaksylykkyzy K., Nabyeva G. Современные методы психолингвистики и нейролингвистики в усвоении второго языка [Electronic resource] // Eurasian Journal of Philology Science and Education/ 2021. Vol. 183, № 3. <https://doi.org/10.26577/EJPh.2021.v183.i3.ph2>
8. Draganski B., May A. Training-induced structural changes in the adult human brain // Behavioral Brain Research. 2008. № 192. P. 137–142.

9. Structural plasticity in the bilingual brain: Proficiency in a second language and age of acquisition affect grey matter density / Mechelli A., Crinion J., Noppeney U., O'Doherty J., Ashburner J., Frackowiak R. // *Nature*. 2004. 757 p.
10. Structural plasticity in the language system related to increased second language proficiency / Stein M., Federspiel A., Koenig T., Wirth M., Lehmann C., Strik W., Wiest R., Brandeis D., Dierks T. // *Cortex*. 2012. № 48. P. 458–465.
11. Green, D. W., Crinion, J., & Price, C. J. Exploring cross-linguistic vocabulary effects on brain structures using voxel-based morphometry. *Bilingualism: Language and Cognition*. 2007. № 10 (2). P. 189–199.
12. Позднякова Е.М. Нейролингвистика в преподавании и изучении иностранных языков. М.: МГИМО, 2023. 20 с.
13. Пуцилло М. В., Винокуров А. Г., Белов А. И. Атлас «Нейрохирургическая анатомия». Под редакцией А. Н. Коновалова. М. : Антидор, 2002. 206 с.
14. Ермолаева Ж. Е., Герасимова И. Н. К вопросу о нейроандрагогике в языковом обучении и консультировании взрослых // *Культура и безопасность*. 2022. № 4. С. 63–70.
15. Писарь Н. В. Потенциал использования нейросетей как инновационного инструмента создания учебного контента и средства организации интерактивной образовательной среды на занятиях по русскому языку как иностранному [Электронный ресурс] // *Филологические науки. Вопросы теории и практики*. 2024. № 17(1). С. 58–65. DOI: <https://doi.org/10.30853/phil20240009/>.

УДК 37.01

Разорвина Анна Сергеевна

кандидат педагогических наук,
доцент кафедры иностранных языков
Ульяновский государственный
педагогический университет
им. И. Н. Ульянова
г. Ульяновск, Россия

Anna Razorvina

PhD in Pedagogy, Associate Professor
of the Department of Foreign
Languages
Ulyanovsk State Pedagogical
University named after I. N. Ulyanov
Ulyanovsk, Russia
anna141281@mail.ru

Куварина Екатерина Николаевна

старший преподаватель кафедры
иностранных языков
Ульяновский государственный
педагогический университет
им. И. Н. Ульянова
г. Ульяновск, Россия

Ekaterina Kuvarina

Senior Lecturer of the Department
of Foreign Languages
Ulyanovsk State Pedagogical
University named after I. N. Ulyanov
Ulyanovsk, Russia
en.kuvarina@gmail.com

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОБРАЗОВАТЕЛЬНЫХ ЦЕЛЯХ: МОТИВАЦИОННЫЕ, ЭТИЧЕСКИЕ И КОГНИТИВНЫЕ АСПЕКТЫ

В данной статье рассматривается проблема восприятия технологии искусственного интеллекта (ИИ) в образовательной среде, с акцентом на мотивационные, когнитивные и этические аспекты. На основе анкетного опроса более 100 респондентов были выявлены различия в отношении к ИИ, уровне цифровой компетенции и мотивации к его использованию. Выделены три группы респондентов: энтузиасты, осторожные пользователи и скептики. Установлено, что среди студентов преобладают энтузиасты, демонстрирующие высокий уровень доверия к ИИ и активное использование творческих инструментов. Преподаватели, напротив, проявляют больше скепсиса, что объясняется возрастным и профессиональным барьером, отсутствием методической базы и этическим опасениями. В статье также предпринята попытка оценить дальнейшую траекторию развития ИИ в образовании. Полученные данные позволяют глубже понять факторы влияния ИИ на образование и подчеркивают необходимость более взвешенного подхода и дифференцированной стратегии интеграции технологии в обучение с учетом характеристик целевых групп.

Ключевые слова: искусственный интеллект (ИИ); иностранный язык; образование; тенденции в образовании; мотивация.

ARTIFICIAL INTELLIGENCE IN EDUCATION: MOTIVATIONAL, ETHICAL, AND COGNITIVE ASPECTS

This article studies the problem of perception of artificial intelligence (AI) technology in the educational environment, with an emphasis on motivational, cognitive and ethical aspects. Based on a questionnaire survey of more than 100 respondents, differences in attitudes towards AI, the level of digital competence and motivation to use it were revealed. There are three groups of respondents: enthusiasts, cautious users, and skeptics. It has been established that enthusiasts

predominate among students, demonstrating a high level of trust in AI and the active use of creative tools. Teachers, on the other hand, show more skepticism. This is due to the age and professional barrier, lack of a methodological framework, and ethical concerns. The article also attempts to assess the further trajectory of AI development in education. The data obtained allow for a deeper understanding of the factors influencing AI on education and emphasize the need for a more thoughtful approach and a differentiated strategy for integrating technology into education, taking into account the characteristics of target groups.

Key words: artificial intelligence (AI); foreign language; education; trends in education; motivation.

Современные ИИ-технологии (нейросетевые модели, NLP, большие языковые модели, такие как GPT) позволяют создавать интерактивные образовательные среды и персонализированные системы обучения. Такие системы могут анализировать огромные объёмы данных об учебном процессе и адаптировать обучение под каждого конкретного студента. Например, платформа Duolingo использует ИИ для построения индивидуальной учебной траектории с геймифицированной подачей материала.

Отечественные и зарубежные исследователи в целом сходятся во взглядах на потенциал ИИ. Российские авторы подчеркивают индивидуализацию обучения и изменение роли учителя. Так, Карпович Э. М. отмечает, что ИИ-технологии могут адаптировать процесс обучения под каждого студента и существенно снизить нагрузку на преподавателя (подготовка к занятиям, проверка работ) [1]. Кувшинова Е. Е. подчеркивает, что на современном этапе развития использование ИИ стало частью нашей жизни, а роль преподавателя трансформируется: от роли источника знаний он переходит к роли тьютора, помощника [2]. Зарубежные исследователи также отмечают положительное влияние ИИ на процесс обучения. Так, например, обзор показал, что ИИ-приложения доказали свою эффективность в улучшении рецептивных и продуктивных языковых навыков, расширении словарного запаса и даже развитии межкультурной компетенции [3]. Анализ исследований подтверждает повышение успеваемости и мотивации при использовании ИИ-инструментов [4]. Вместе с тем и отечественные, и зарубежные авторы обращают внимание на вызовы, которые несут в себе ИИ-технологии: технические особенности, этические вопросы, педагогические и психологические проблемы [1; 4].

Целью нашего исследования стало изучение вовлеченности и общего настроения студентов по отношению к технологиям, использующим искусственный интеллект. В исследовании приняли участие 89 студентов ФГБОУ ВО «УлГПУ им. И. Н. Ульянова». Цель опроса – определить уровень вовлеченности в практику использования ИИ, выявить доминирующие установки и возможные опасения, а также отношение к перспективам использования искусственного интеллекта в образовательной среде.

На основании анкетных данных нами были выделены три группы респондентов, в зависимости от их отношения к ИИ, мотивации и уровня цифровой компетенции. Полученные данные представлены в Таблице 1.

Таблица 1

Распределение студентов по группам

Группа	Количество студентов	Доля от выборки
Энтузиасты	41	46 %
Осторожные	33	37 %
Скептики	15	17 %
Всего	89	100 %

1. Энтузиасты (n=41, 46 %)

Критерии, по которым респондентов отнесли в данную группу: положительное отношение к технологии ИИ, ее понимание и положительное отношение к ИИ в образовании.

Данная группа характеризуется высоким уровнем доверия к технологиям, у них четко выражено ощущение пользы ИИ: 83,7 % считают, что он значительно или умеренно улучшает образование. Респонденты данной группы высказывают меньше опасений по сравнению с другими группами, поддерживают использование ИИ в творчестве (34 % – «поддерживаю с оговорками», 20 % – «Полностью поддерживаю»). Однако, они реже других считают, что ИИ может заменить преподавателей. Студенты данной группы чаще других отмечали использование творческих инструментов ИИ – генераторы текстов, идей, изображений и презентаций, а также отмечалась и готовность использовать ИИ вне академических задач.

2. Осторожные пользователи (n=33, 37 %)

В данную группу вошли студенты, которые используют ИИ с опасением или считают свое знание технологий ИИ удовлетворительным, в целом не проявляют отрицания, но испытывают сомнения.

Такие студенты демонстрируют умеренное отношение к ИИ, большинство считает, что он немного улучшает образование. Однако, у данной группы опасения выражены четче: преобладают страхи неэтичного использования, ошибочных решений и возможной зависимости от технологии, многие выбирают нейтральную позицию относительно ИИ в творчестве. Обучающиеся из данной группы чаще используют переводчики и инструменты проверки ошибок, но отмечают тот факт, что они избегают использовать генератор текстов.

3. Скептики (n=15, 17 %)

Студенты рассматриваемой группы отмечают, что им не нравится использовать ИИ, но вынуждены это делать, отсутствует уверенность в эффективности данной технологии.

В их настрое относительно изучаемой технологии просматривается недоверие: они считают, что ИИ не способен или маловероятно способен

повысить качество образования, выражают максимальные опасения, включая потерю рабочих мест, этические дилеммы и замещение преподавателей. В сфере творчества они видят угрозу подлинному авторству: чаще выбирают позиции «против». Чаще всего отмечают использование проверок текстов на плагиат и использование поисковых систем.

Типология студентов отражает разнообразие подходов и стратегий по отношению к искусственному интеллекту в образовательном процессе. Интересно, что несмотря на широкое распространение данной технологии, в группу энтузиастов вошло менее половины всех респондентов. Осторожные составляют промежуточную группу и могут склониться как в сторону доверия, так и усиления недоверия, в зависимости от образовательной политики.

В продолжение исследования мы сравнили мнение студентов и педагогов по поводу технологии ИИ и ее потенциала. Так, среди педагогов, выборка была незначительной, однако нам также удалось выделить три группы: энтузиасты, осторожные и скептики.

1. Энтузиасты отличаются открытостью к цифровым инновациям, видят в ИИ не угрозу, а ресурс для повышения интереса к обучению, адаптации заданий и развития цифровых компетенций студентов. Такие педагоги готовы применять ИИ в проектной и визуальной деятельности.

2. Осторожные пользователи зачастую не отрицают полезность ИИ, но подчеркивают ограничения: недостаточную цифровую грамотность студентов, слабое методическое сопровождение. Такие педагоги чаще используют ИИ как вспомогательный ресурс, без глубокого включения его в учебный процесс.

3. Скептики выражают обеспокоенность неограниченным использованием технологий, утратой личного вклада. Такие преподаватели подчеркивают, что ИИ не может заменить живое общение, а в некоторых случаях даже мешает.

В сравнении со студентами, педагоги распределились несколько иначе. Результаты представлены в Таблице 2.

Таблица 2
Сравнение педагогов и студентов

Группа	Преподаватели	Студенты
Энтузиасты	36 %	46 %
Осторожные	36 %	37 %
Скептики	28 %	17 %
Всего	100 %	100 %

В целом, анализ показывает, что среди преподавателей больше скептиков, а доля энтузиастов ниже, что обусловлено рядом факторов. Во-первых,

возрастной и технологический барьер, свойственный старшему поколению педагогов, который проявляется в менее уверенном пользовании цифровыми инструментами и сдержанностью в их применении. Во-вторых, отсутствие нормативно-методической базы по внедрению ИИ в образовательный процесс, что порождает настроенность в использовании данных технологий со стороны педагогов. В-третьих, преподаватели выражают опасения, связанные с подменой самостоятельной когнитивной творческой деятельности современными технологиями, что противоречит целям традиционного образования. Четвертым фактором, на наш взгляд, выступает угроза подрыва профессиональной идентичности педагога на фоне развития интеллектуальных систем. И наконец, различия могут быть объяснены и функциональной мотивацией: если студенты используют ИИ преимущественно для упрощения и ускорения учебных задач, то преподаватели оценивают его через призму педагогической эффективности, зачастую не находя убедительных доказательств.

На наш взгляд, перспективы развития ИИ в обучении будут развиваться по следующим направлениям:

Создание эмоционально-чувствительных ИИ-ассистентов, которые способны анализировать не только языковую подготовку и уровень обучающихся, но и их эмоциональное состояние, мотивационные установки и динамику утомления. Такие системы смогут подирать содержание и тип взаимодействия в зависимости от когнитивного и аффективного состояния ученика, выступая в роли наставника, ориентированного на индивидуальный психоэмоциональный профиль. Данное направление пока представлено в рамках адаптивного образования [5], однако, требует дальнейшего изучения и междисциплинарных подходов, объединяющих лингводидактику, нейропсихологию и технологии эмоционального искусственного интеллекта.

Создание концепции «языковой памяти» – персонализированного ИИ-профиля обучающегося, который запоминает всю динамику и историю ошибок, успешных стратегий обучения именно для данного студента на протяжении всего образовательного пути. Данная система может также стать основой для построения адаптивной траектории обучения. При чем ИИ будет не просто справлять ошибки, но и обучать выразительности речи и структурированию текста, учитывая особенности обучающегося.

Создание ИИ-обучения через призму социокультурной компетенции. В настоящее время широко обсуждается данное направление [3; 4], но в отличие от существующих решений, мы предполагаем, что необходимо расширить возможности ИИ для моделирования речевых стратегий с учетом менталитета обучающегося и социокультурного контекста изучаемого языка. Это позволит формировать речевое поведение, характерное коммуникативным нормам целевой культуры, не нарушая при этом культурной иден-

тичности самого студента. В такой ситуации ИИ становится проводником в социокультурную среду языка, что делает его инструментом формирования подлинной межкультурной компетенции.

Таким образом, проведенный опрос выявил расслоение в восприятии и использовании ИИ в образовательной среде как среди студентов, так и среди преподавателей. Полученные результаты подчеркивают важность учета мотивационных и когнитивных различий между целевыми группами при интеграции ИИ в учебный процесс. Необходимо отметить, что несмотря на высокую технологическую осведомленность и готовность респондентов, наличие скепсиса и настороженности свидетельствует о необходимости комплексной поддержки: нормативной и педагогической, формирование культуры осознанного и критического использования данных технологии.

ЛИТЕРАТУРА

1. Карпович Э. М. Применение искусственного интеллекта в обучении иностранным языкам // Филологический аспект: международный научно-практический журнал. Сер.: Методика преподавания языка и литературы. 2023. № 04 (21). URL: <https://scipress.ru/fam/articles/primenenie-iskusstvennogo-intellekta-v-obuchenii-inostrannym-yazykam.html> (дата обращения: 14.07.2025).
2. Кувшинова Е. Е. Применение искусственного интеллекта в обучении иностранному языку // Гуманитарий Юга России. 2024. Том 13. № 2. С. 75–84. DOI: <https://doi.org/10.18522/2227-8656.2024.2.7>.
3. AlTwijri A. A., Alduais A. M. S. Artificial intelligence (AI) and English language learning: A systematic review // Heliyon. 2023. Vol. 9, Issue 5. e15499. DOI: <https://doi.org/10.1016/j.heliyon.2023.e15499>.
4. Artificial intelligence in education: A systematic literature review / Shan W., Fang W., Zhen Zh., Jingxuan W., Tam T., Zhao D. // Expert Systems with Applications. 2024. Vol. 252. Part A. 124167. DOI: <https://doi.org/10.1016/j.eswa.2024.124167>.
5. Алферьева А. А. Искусственный интеллект в образовании: как адаптивное обучение и цифровые ассистенты меняют подход к обучению и воспитанию подростков // Вестник науки №1 (82). 2025. том 1. С. 111–119. URL: <https://www.вестник-науки.pf/article/20559> (дата обращения: 14.07.2025).

УДК 372.881.111.1

Руденко Татьяна Николаевна
кандидат филологических наук,
доцент, доцент кафедры
лексикологии и стилистики
английского языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Tatiana Roudenko
PhD in Philology, Associate Professor,
Associate Professor of the Department
of Lexicology and Stylistics
of the English Language
Belarusian State University
of Foreign Languages
Minsk, Belarus
tatianaroudenko@gmail.com

ИНТЕГРАЦИЯ ЦИФРОВЫХ ТЕХНОЛОГИЙ ДЛЯ РАЗВИТИЯ МЕЖКУЛЬТУРНОЙ КОМПЕТЕНТНОСТИ

Статья посвящена анализу возможностей интеграции искусственного интеллекта, виртуальной и дополненной реальности, а также платформ цифрового сотрудничества в курс «Язык средств массовой информации», который занимает центральное место в лингвистической подготовке по направлению 1-21 06 01-01 «Современные иностранные языки (преподавание)». Предложенные подходы способствуют углублению понимания лексических, грамматических и жанрово-стилистических особенностей, семиотических ресурсов и конвенций современного дискурса, раскрывая механизмы конструирования смыслов в СМИ, способы «оформления» событий и влияние на общественное восприятие в разных культурных контекстах. В качестве основных инструментов рассматриваются симуляции на базе ИИ, интерактивные сценарии VR/AR и структурированные обмены на платформах цифрового сотрудничества. Необходимым условием для освоения цифровых инструментов и формирования межкультурной компетентности становится сочетание технологических инноваций и обоснованных этических принципов.

Ключевые слова: искусственный интеллект; виртуальная и дополненная реальность; платформы цифрового сотрудничества; цифровая грамотность; межкультурная компетентность.

INTEGRATING DIGITAL TECHNOLOGIES FOR INTERCULTURAL COMPETENCE

The article examines ways of integration of artificial intelligence, virtual and augmented reality, telecollaboration platforms, and mobile applications into the “Language of Mass Media” course for future ESL teachers which can operationalize five learning outcomes – mastery of journalistic English, enhanced intercultural communication, rigorous media analysis, contextualized socio-political interpretation, and cultivation of civic engagement – through AI-driven simulations, immersive VR/AR scenarios, structured eTandem exchanges, and multimedia mobile tasks. The study addresses implementation challenges – digital equity, instructor readiness, ethical data practices, and cultural authenticity – and proposes mitigation via hybrid models, targeted professional development, and human-in-the-loop oversight. Future research directions include longitudinal assessments of AI-mediated competence, bias-aware ethical design, AR-enhanced storytelling, and pedagogy that harmonizes real-time translation

with intercultural nuance training. By synthesizing technological innovation with robust pedagogical and ethical principles, the paper offers a strategic blueprint for equipping ESL educators with the digital fluency and intercultural agility required in today's media-driven world.

Key words: artificial intelligence; virtual and augmented reality; telecollaboration platforms; digital literacies; intercultural competence.

It is widely acknowledged that the rapid proliferation of digital technologies has fundamentally reshaped language teacher education by enriching its intercultural dimension and fostering critical literacies. Integrating tools such as artificial intelligence, virtual and augmented reality, and telecollaboration platforms enables prospective foreign-language instructors to develop digital fluency, critical thinking, and intercultural competence in concert. Anchored in this paradigm, the course “Language of Mass Media” occupies a pivotal place in the linguistic training of future ESL teachers. As its title indicates, the course examines the linguistic frameworks, semiotic resources, and discursive conventions that underpin contemporary news discourse, revealing how media texts construct meaning, frame events, and influence public perception across cultural contexts.

Designed to cultivate advanced communicative and analytical skills, “Language of Mass Media” pursues five core objectives. First, it deepens students' understanding of the lexical, grammatical, and genre-stylistic features characteristic of journalistic English. Second, it enhances communicative competence in intercultural settings by engaging learners in authentic news-based tasks. Third, it cultivates rigorous methodologies for linguistic analysis, enabling trainees to dissect English-language media texts with precision. Fourth, it strengthens abilities to interpret socio-political and socio-cultural phenomena within British and American contexts – situated historically and compared with Belarusian realities. Finally, it fosters the formation of socio-cultural values, civic engagement, and political awareness. By weaving digital technologies into each module – through AI-driven content simulations, VR newsroom immersions, and telecollaborative projects – the course not only aligns with globalized communication demands but also equips future educators with the critical insights and practical tools required to navigate, analyze, and teach within today's dynamic media landscape. This course affords unparalleled opportunities to integrate contemporary digital technologies into the cultivation of intercultural competence among future ESL teachers; its specific applications are detailed below.

A coherent approach to digital integration rests on established models of intercultural competence and technology-enhanced learning. Key frameworks include Deardorff's Process Model of Intercultural Competence, emphasizing attitude, knowledge, skills, and reflection [2]; TPACK (Technological Pedagogical Content Knowledge), which guides teachers in blending content mastery with digital tools; Digital Literacy Continuum, outlining progressive competences

from operational to critical evaluation of media. Aligning these models allows curriculum designers to scaffold digital tasks that advance both intercultural sensitivity and pedagogical expertise.

Digital innovations have fundamentally transformed the methodologies through which language educators cultivate learners' intercultural competence. Four principal technological categories emerge as especially influential: Artificial Intelligence (AI), Virtual Reality (VR) & Augmented Reality (AR), telecollaboration platforms, and mobile & social media applications. AI-driven tools employ conversational agents, adaptive feedback algorithms, and scenario generators to simulate authentic dialogues, personalize learning trajectories, and foster reflective intercultural analysis. Similarly, VR & AR environments create immersive cultural simulations – ranging from virtual marketplaces to formal negotiation settings – that situate learners within lifelike contexts, thereby deepening empathic engagement and contextual understanding. Telecollaboration platforms, including eTandem exchanges and synchronous video-conference rooms, enable real-time interaction with peers across geopolitical boundaries, promoting pragmatic competence and cross-cultural dialogue. Finally, mobile and social media applications facilitate multimedia storytelling, micro-learning modules, and real-time polling, offering flexible, learner-centered opportunities to document cultural observations and receive immediate feedback. As these technologies evolve, educators must critically assess issues of access, ethical data use, and cultural authenticity to ensure that pedagogical designs remain equitable and deeply rooted in real-world intercultural dynamics [1].

Artificial intelligence-enhanced applications offer unprecedented opportunities to tailor and amplify intercultural learning experiences. Chatbot interfaces, endowed with distinct cultural personas, enable learners to engage in iterative role-plays that mirror both casual and professional interactions, thus fostering pragmatic fluency within a safe, repeatable environment. Complementing these simulations, machine-learning algorithms dynamically calibrate task complexity – modulating lexical difficulty, phonetic nuance, and discourse register in response to individual performance data. Concurrently, automated e-journals leverage natural language processing to dissect learners' written reflections, identifying cultural markers, speech acts, and pragmatic choices that reveal developing intercultural sensitivity. This analytical feedback loop not only deepens metacognitive awareness but also generates actionable insights for instructors. Crucially, the pedagogical integration of AI must remain under vigilant human supervision: educators review algorithmic recommendations, contextualize corrective prompts, and guard against potential biases embedded in the training data. By harmonizing AI's scalability with teacher expertise, language programs can deliver culturally attuned instruction that is both equitable and reflective of complex real-world dynamics.

Virtual and augmented reality technologies enable learners to inhabit nuanced cultural habitats. Immersive platforms such as ImmerseMe and Mondly VR reconstruct authentic microspaces – from the terraces of Parisian cafés to the alleys of Tokyo’s fish markets – providing richly contextualized scenarios for pragmatic exploration. Within these environments, students engage in negotiation simulations, etiquette walkthroughs, and context-driven problem-solving, negotiating social cues and nonverbal norms in real time. This embodied approach enhances cognitive presence and emotional resonance, stimulating deeper intercultural empathy and narrative immersion. Nevertheless, implementation challenges persist: high initial investment in headsets and software, ongoing maintenance demands, and inequitable access across socio-economic contexts. To democratize immersive learning, educators should adopt hybrid pedagogical models that integrate VR modules with accessible AR overlays on mobile devices and complementary low-tech role-plays. Rigorous evaluation of learning outcomes, coupled with qualitative studies on learner presence and cultural appropriation, will guide the creation of equitable, scalable immersive curricula.

Telecollaboration platforms create authentic intercultural exchanges by linking ESL learners with native-speaker peers in structured, pedagogically scaffolded environments. In eTandem partnerships, dyads negotiate linguistic and cultural nuances through guided tasks, alternating roles as tutor and learner to deepen mutual understanding. Synchronous group projects conducted via Zoom, Microsoft Teams, or bespoke educational networks allow small teams to co-construct presentations, conduct surveys, or solve case studies, fostering real-time negotiation of meaning and shared problem-solving across time zones. Asynchronous collaboration thrives in digital portfolios, where participants document cultural observations, post reflective commentaries, and critique one another’s entries, promoting metacognitive awareness and peer-mediated feedback. Despite their pedagogical promise, these technologies demand careful coordination: scheduling must reconcile disparate global calendars, connectivity must be robust enough to support seamless interaction, and participants require sufficient digital literacy to navigate multiple interfaces. To maximize efficacy, instructors should integrate clear orientation modules, establish regular check-ins, and provide contingency plans – such as low-bandwidth alternatives or flexible deadlines – to ensure equitable access and sustained engagement [4].

In contemporary second-language instruction, mobile devices facilitate learner-centered intercultural engagement across multiple modalities. Multimedia storytelling applications enable students to craft digital narratives that foreground diverse cultural perspectives, integrating text, audio, and visual elements to foster narrative empathy and critical awareness. Simultaneously, microlearning modules – delivered via mobile platforms – present concise cultural insights through interactive quizzes, infographics, and brief video clips, supporting just-in-time knowledge acquisition and scaffolded reflection. Real-time polling and

threaded discussions further provide instantaneous feedback loops, cultivating collaborative examination of intercultural themes and promoting dynamic peer exchange, both asynchronously and during synchronous seminar sessions.

However, effective deployment of such technologies demands careful design considerations. Educators and developers must ensure broad device compatibility to accommodate varied hardware and operating systems while optimizing multimedia content for minimal data consumption. They should implement curated interfaces and establish clear usage protocols to mitigate distractions endemic to social media environments. Additionally, pedagogical frameworks require explicit scaffolding to guide learners through mobile-mediated cultural inquiry, ensuring that digital tools align seamlessly with established curricular objectives and desired learning outcomes.

In ESL/Intercultural Training technologies are integrated through structured design. In contemporary ESL curricula, task-based learning fosters intercultural competence through outcome-driven digital projects. Educators design authentic assignments – such as virtual intercultural interviews – that require learners to negotiate linguistic nuances and cultural frameworks while collaborating with global peers. Scaffolding precedes performance, with pre-brief sessions outlining relevant cultural norms and communication protocols. After task completion, structured reflection is facilitated through e-journals or AI-powered analytics, prompting metacognitive appraisal of discourse strategies and cultural interpretations. Assessment employs multifaceted rubrics that synthesize AI-generated discourse analyses – highlighting pragmatic markers and interactional features – with instructor evaluations to ensure nuanced feedback on both linguistic performance and intercultural sensitivity. A blended pedagogical approach further amplifies learning transfer by interweaving traditional face-to-face role-plays with immersive digital simulations, enabling cyclical practice across modalities. This integration of targeted tasks, reflective scaffolds, robust assessment, and methodological pluralism equips learners with adaptable intercultural strategies for real-world communication [3].

Designing an integrative evaluation model necessitates multidimensional tools that reflect learners' progression in both digital literacies and intercultural awareness. Digital portfolios serve as curated repositories – comprising recorded dialogues, reflective essays, and multimedia presentations – that collectively map trajectories in cultural understanding and technological dexterity. Analytics dashboards complement these portfolios by offering visualizations of engagement patterns, linguistic accuracy, and the sophistication of reflective entries, enabling instructors to identify emerging trends and tailor pedagogical interventions. Rubric-driven reviews then furnish transparent criteria for evaluating intercultural sensitivity, communicative strategies, and proficiency with digital platforms. Finally, peer and self-assessment mechanisms provide scaffolded opportunities for learners to engage in guided critique of their own and their peers' intercultural

performances, thereby reinforcing metacognitive skills and community-based reflection. Through this multifaceted framework, robust evaluation ensures congruence between intended learning outcomes and the affordances of contemporary educational technologies.

Despite the promise of digital tools to democratize intercultural learning, a complex array of structural and pedagogical challenges must be addressed. First, the digital divide persists: unequal access to reliable devices and broadband networks limits learners' ability to engage with VR simulations, telecollaboration exchanges, or AI-driven platforms, reinforcing existing educational inequities. Second, teacher digital literacy emerges as a pivotal factor; ongoing professional development is essential to equip instructors with the skills to integrate novel technologies seamlessly and to troubleshoot technical issues in real time. Third, ethical and privacy concerns cannot be overlooked – robust data-security protocols, transparent informed-consent procedures, and clear policies on AI-collected interaction logs are vital to protect learner confidentiality and autonomy. Finally, cultural accuracy in AI-mediated environments demands vigilant human oversight: algorithmic training data may inadvertently perpetuate stereotypes or flatten cultural nuance, so educators must curate content, contextualize machine outputs, and intervene when representations risk oversimplifying complex sociocultural realities. By confronting these interrelated challenges through policy, pedagogy, and ethical praxis, language programs can harness technology to foster equitable, authentic, and culturally attuned communication.

Emerging research trajectories in digital intercultural pedagogy underscore the necessity for longitudinal investigations into AI-mediated intercultural competence development. By systematically tracking learner cohorts across multiple educational settings, researchers can discern the enduring effects of adaptive algorithms on pragmatic sensitivity, strategic communicative choices, and cultural metacognition. Concurrently, the formulation of robust ethical frameworks is imperative to guide culturally responsive AI design in language education – prioritizing transparency, bias mitigation, and learner autonomy while respecting the diversity of sociocultural narratives. Furthermore, augmented reality-enhanced storytelling emerges as a fertile ground for deepening narrative empathy: learners not only consume but actively co-create culturally situated stories, navigating immersive environments that integrate multimodal artifacts, social context, and symbolic cues. Finally, the integration of real-time translation tools must be thoughtfully calibrated with explicit intercultural nuance training; pairing instantaneous linguistic mediation with pedagogical interventions that challenge reductive equivalences and cultivate critical awareness of sociolinguistic registers [6]. Future research should adopt an interdisciplinary approach – bridging applied linguistics, instructional design, data ethics, and human-computer interaction – to construct an inclusive, ethically grounded, and technologically sophisticated agenda for advancing intercultural language instruction [5].

Beyond the four core technologies, a suite of emerging digital tools can significantly deepen intercultural ESL instruction. Digital storytelling platforms such as Adobe Spark, Storybird, and WeVideo enable learners to co-author multimodal narratives that integrate text, audio, video, and interactive maps. For instance, StoryMapJS and ArcGIS StoryMaps allow students to situate cultural narratives within geospatial frameworks, fostering place-based empathy.

AI-driven pronunciation and accent trainers – including ELSA Speak, Speechace, and PronunciationCoach – offer fine-grained phonetic feedback, pitch-contour visualizations, and adaptive practice sequences to support authentic intonation and rhythm. Complementing these, AI-mediated chat simulators (e.g., Replika, ChatGPT with custom cultural personas) provide safe environments for iterative intercultural dialogues in business, academic, or social contexts.

Blockchain-enabled e-portfolios built on platforms like Blockcerts, Canvas Badges, and Mahara-Hyperledger ensure tamper-proof documentation of critical intercultural milestones, digital badges, and microcredentials. These secure records enhance learner ownership and support longitudinal tracking of reflective practices.

Additional innovations – social annotation tools (Hypothesis) for collaborative discourse analysis, AR translation apps (Google Lens, Microsoft Translator) for on-the-spot cultural mediation, and mobile AR platforms (HP Reveal, Metaverse) for contextualized site-based learning – further enrich task-based and blended activities. Integrating this expanded toolkit into project-driven modules (such as VR ethnographic interviews, blockchain-anchored glossaries, and collaborative digital dossiers) strengthens the continuum of immersive practice, iterative reflection, and comprehensive assessment across the digital-intercultural learning spectrum.

Equally vital is the alignment of new technologies with robust theoretical and ethical frameworks. Advancing cultural intelligence (CQ) models through interdisciplinary collaboration – partnering with computer-science experts to design bias-aware algorithms – will help trainees develop culturally responsive, human-centric learning environments. Embedding these frameworks within mixed-method evaluation protocols ensures that AI-generated analytics and human judgment co-here, safeguarding authenticity and equity. By weaving these innovations into the pedagogical tapestry of AI chatbots, VR/AR simulations, telecollaboration, and mobile tools, ESL educators can deliver holistic, technologically advanced programs that cultivate both linguistic proficiency and intercultural agility. Beyond these core areas, future educators might explore blockchain for secure e-portfolios, AI-driven accent trainers, and multimodal digital storytelling platforms. Integrating cultural intelligence frameworks with advanced technologies will further enrich ESL teacher preparation.

In conclusion, the systematic integration of artificial intelligence, virtual and augmented reality, telecollaboration platforms, and mobile applications into the

“Language of Mass Media” curriculum – underpinned by Deardorff’s Process Model of Intercultural Competence and the TPACK framework – serves to develop future ESL teachers’ digital literacies, intercultural proficiency, and ethical discernment, thereby positioning them to meet the exigencies of contemporary media environments and to enact culturally responsive pedagogy.

LITERATURE

1. AI in Cultural Contexts: Teaching Intercultural Communication to EFL/ESL Students [Electronic resource] // EFLCafe.net. URL: <https://eflcafe.net/ai-in-cultural-contexts-teaching-intercultural-communication-to-efl-esl-students/> (date of access: 01.07.2025).
2. Deardorff D.K. Manual for Developing Intercultural Competencies: Story Circles [Electronic resource] // Routledge Focus. London, 2019. 101 p. DOI: <https://doi.org/10.4324/9780429244612>.
3. ESOL Teaching Strategies [Electronic resource] // LATs Website. URL: [https://activitytypes.wm.edu/ESOLTeachingStrategies-August2012\(2\).htm](https://activitytypes.wm.edu/ESOLTeachingStrategies-August2012(2).htm) (date of access: 01.07.2025).
4. eTandem Learning: Autonomous Language Learning with a Partner [Electronic resource] // Pools of Open Online Learning Systems. URL: <https://www.languages.dk/archive/Methods/manuals/E-Tandem/e-tandemUK.pdf> (date of access: 01.07.2025).
5. Klimova B. The Impact of AI on Enhancing Students’ Intercultural Communication Competence at the University Level / B. Klimova, J.H. Chen [Electronic resource] // Language Teaching Research Quarterly. 2024. Vol. 43, P. 102–120. URL: <https://api.eurokd.com/Uploads/Article/1267/ltrq.2024.43.06.pdf> (date of access: 01.07.2025).
6. Song X. Using Cross-Cultural Machine Translation Technology to Promote Communication and Cooperation in English Courses / X.Song, Y.Cheng, T.Zeng [Electronic resource] // Journal of Electrical Systems. 2024. Vol. 20, P. 412–418. DOI: <https://doi.org/10.52783/jes.2665>.

УДК 378.147

Савчик Ксения Игоревна

учитель высшей категории,
учитель немецкого языка
ГУО «Средняя школа № 28 имени
В. Д. Соколовского г. Гродно»
г. Гродно, Беларусь

Kseniya Sauchyк

teacher of the highest category, teacher
of German
SEI «Secondary school № 28
Grodno named after W.D Sokolovski»
Grodno, Belarus
ksyusha-savchik@mail.ru

ФОКУС НА ПРЕИМУЩЕСТВАХ И НЕДОСТАТКАХ ЦИФРОВОГО ОБУЧЕНИЯ

Цифровое обучение в последние годы стремительно развивается, и его влияние на образовательный процесс становится все более значимым. Данная статья рассматривает цифровое обучение как сложный и многогранный феномен, который требует внимательного анализа и взвешенного подхода. Цифровое обучение предоставляет доступ к огромному количеству ресурсов и материалов, что делает образование более гибким и разнообразным. В статье представлены преимущества и недостатки цифрового обучения, характеристики цифровых инструментов, которые предоставляют огромные возможности для обучения и самовыражения. Для успешной интеграции цифровых технологий в образовательный процесс необходимо учитывать, как педагогические, так и социальные аспекты, направленные на минимизацию рисков и максимизацию возможностей. Использование цифровых инструментов требует не только технической грамотности, но и умения критически оценивать информацию, умения адаптировать методы преподавания к новым условиям. Без должной подготовки и необходимых знаний использование цифровых технологий может привести к снижению качества образования. В данной статье акцентируется внимание также на потенциале искусственного интеллекта, который представляет собой мощный инструмент, способный трансформировать образовательный ландшафт. Технологии искусственного интеллекта становятся все доступнее, открывают новые горизонты для обучения и преподавания. Искусственный интеллект может не только автоматизировать рутинные задачи, но и предлагает персонализированные пути обучения, адаптируется к потребностям каждого участника образовательного процесса.

Ключевые слова: цифровизация; цифровые инструменты; визуализация; медиаграмотность; искусственный интеллект.

FOCUS ON THE ADVANTAGES AND DISADVANTAGES OF DIGITAL LEARNING

Digital learning has been rapidly developing in recent years, and its impact on the educational process is becoming increasingly significant. This article examines digital learning as a complex and multifaceted phenomenon that requires careful analysis and a balanced approach. Digital learning provides access to a huge number of resources and materials, which makes education more flexible and diverse. The article presents the advantages and disadvantages of digital learning, the characteristics of digital tools that provide enormous opportunities for learning and self-expression. For the successful integration of digital technologies into the educational process, it is necessary to take into account both pedagogical and social aspects aimed at minimizing risks and maximizing opportunities. The use of digital tools requires not only technical literacy, but also the ability to appreciate critically information,

the ability to adapt teaching methods to new conditions. Without proper training and necessary knowledge, the use of digital technologies can lead to a decrease in the quality of education. This article also focuses on the potential of artificial intelligence, which is a powerful tool that can transform the educational landscape. Artificial intelligence technologies are becoming more accessible, opening up new horizons for learning and teaching. Artificial intelligence can not only automate routine tasks, but also offer personalized learning paths, adapting to the needs of each participant in the educational process.

Key words: digitalization; digital tools; visualization; media literacy; artificial intelligence.

Цифровое обучение стало неотъемлемой частью системы образования. Цифровое обучение подразумевает собой замену печатного контента на цифровые носители, что позволяет легко передавать знания. Используя цифровые инструменты, педагоги могут преподносить контент таким образом, чтобы привлечь внимание учащихся и повысить их интерес к изучению языка.

Цифровые инструменты позволяют быстро и гибко адаптировать учебные материалы и, таким образом, цифровые учебные материалы имеют явное преимущество, заключающееся в том, что их можно быстро обновлять или добавлять по сравнению с книгами или рабочими листами.

Цифровые презентации обеспечивают динамичный и интерактивный процесс обучения, которые дополняют и обогащают традиционные методы обучения.

Визуализация с помощью цифровых инструментов революционизирует обучение иностранным языкам. С помощью изображений, видео и инфографики можно передавать абстрактные грамматические структуры и лексику таким образом, чтобы они запомнились и были понятны учащимся. Эти наглядные пособия поддерживают процесс обучения, устанавливая прямую связь между новыми знаниями и их практическим применением в реальных условиях. Это делает изучение языка не только более эффективным, но и более увлекательным. Педагоги, использующие цифровые инструменты, не только готовят своих учеников к глобально связанному миру, но и развивают критическое мышление, креативность и цифровую грамотность. Таким образом, интеграция цифровых инструментов в обучение иностранным языкам является важным шагом на пути к современному, интерактивному и ориентированному на будущее подходу к образованию [1].

Цифровыми инструментами обучения являются, например, виртуальные классы, в которых можно не только учиться независимо от местоположения, но и гибко управлять временем. Учащиеся, которые зависят от инвалидной коляски, могут комфортно учиться дома и не находятся в невыгодном положении по сравнению с другими учащимися.

Поскольку у каждого типа учащихся свой темп обучения, еще одним преимуществом цифровых инструментов является независимость от времени. Учащиеся могут получать контент в удобное для них время. При этом есть возможность повторить содержание в любое время, в отличие от фиксиро-

ванных уроков. Такая гибкость поддерживает индивидуальные траектории обучения и помогает учащимся расширять свои знания независимо от временных ограничений.

Приложения виртуальной и дополненной реальности обеспечивают более тесное взаимодействие с учебным предметом, иллюстрируя содержание во время передачи знаний.

Электронные книги в качестве цифровых учебных материалов также могут способствовать цифровому обучению. Большая часть специализированной литературы доступна в онлайн-библиотеках в различных цифровых форматах.

Игровой подход к передаче знаний также очень полезен в образовании. Во время геймификации знания передаются через игры и головоломки. Это не только обучает новому контенту, но и повышает медиаграмотность. Развитие медиаграмотности является важным аспектом, позволяет критически оценивать информацию и эффективно использовать цифровые инструменты для себя и собственного образования.

Целенаправленное использование цифровых обучающих инструментов в обучении охватывает не только умение работать с техническими устройствами, но и развивает важные цифровые навыки.

Таким образом, учащиеся узнают, как распознавать достоверную информацию, как работать с контентом, связанным с защитой данных.

Существует несколько критериев, которым в идеале должен соответствовать носитель информации, чтобы эффективно способствовать процессу обучения:

- индивидуализация учебного контента;
- обучение независимо от времени и места;
- обмен информацией в классе (под руководством учителя);
- воздействие на несколько органов чувств (особенно на зрительные и слуховые);
- расширенные возможности для самостоятельного обучения.

Цифровые платформы и методы обучения открывают беспрецедентные возможности для учащихся и преподавателей, но также сопряжены с определенными проблемами. Как и любая инновация, цифровое обучение имеет не только преимущества, но и недостатки.

Это может быть особенно проблематично для учащихся младшего возраста, для которых развитие социальных навыков является важной частью обучения.

Цифровая среда богата отвлекающими факторами, которые могут мешать процессу обучения. Кроме того, огромный объем доступной информации может ошеломить учащихся и привести к поверхностному взаимодействию с учебным контентом.

Качество цифровых учебных материалов может существенно различаться, и не весь контент одинаково надежен и имеет образовательную ценность. Перед преподавателями стоит задача поиска и использования высококачественных и аутентичных ресурсов.

Также необходимо соблюдать законодательные требования, например, в отношении защиты данных и авторского права.

Чтобы получить все преимущества и свести к минимуму недостатки, решающее значение имеет сбалансированный подход. Это требует тщательного выбора и использования цифровых платформ обучения, учета индивидуальных потребностей в обучении, которые позволяют учащимся критически использовать цифровой контент.

Негативный опыт цифровизации процесса обучения отмечен в Швеции. Шведские школы, которые в течение многих лет были пионерами в использовании цифровых инструментов, оснастили заново начальные школы книгами. В Германии ученые хотят вообще прекратить цифровизацию в школах. Они пришли к выводу, что использование цифровых инструментов в обучении не способствует повышению качества образовательного процесса.

С нейробиологической точки зрения, цифровые инструменты способствуют общему процессу обучения учащихся. Для того, чтобы охватить как можно больше органов чувств, обучение должно быть разнообразным, поскольку мозг состоит из разных структур и областей, некоторые из которых у каждого человека работают по-разному [3].

Современные образовательные подходы признают, что использование цифровых инструментов в классе может обогатить процесс обучения и значительно повысить качество преподавания. Продвижение цифровизации в секторе образования открывает новые возможности для преподавания языков. Для того чтобы цифровые инструменты оказывали положительное воздействие, необходимо отбирать их в соответствии с определенными критериями:

- определение целей учебного занятия;
- ознакомление и соотнесение с техническими возможностями цифровых носителей;
- выбор цифровых инструментов;
- учет аспектов защиты данных и авторского права.

При использовании цифровых инструментов в обучении иностранному языку важно уделять особое внимание методом обучения. Выбор цифровых инструментов для использования в классах должен основываться на возрастных особенностях учащихся, а также на характере учебного материала и целей обучения. Важно учитывать, что разные группы учащихся могут иметь различные потребности и предпочтения в использовании информационных технологий. Если учащиеся уже имеют опыт работы с определенными инструментами, то это значительно ускоряет образовательный процесс. В то же время, если учащиеся сталкиваются с трудностями в использовании цифровых ресурсов, это может негативно сказаться на их обучении и мотивации к их использованию в дальнейшем.

Также стоит акцентировать внимание на необходимости подготовки педагогов к работе в цифровой среде. Преподаватели должны быть не только специалистами в своей предметной области, но и обладать навыками работы

с современными образовательными технологиями. Это включает в себя умение создавать интерактивные материалы, использовать платформы для дистанционного обучения и эффективно взаимодействовать с учениками в онлайн-формате.

В последние годы искусственный интеллект стремительно развивается и применяется в образовании. Все больше преподавателей используют возможности искусственного интеллекта для формирования различных подходов к изучению иностранных языков.

К преимуществам искусственного интеллекта относятся большая скорость и точность выполнения задания. Кроме того, искусственный интеллект снижает вероятность ошибок благодаря высокой точности выполнения задач [2].

Голосовые помощники, такие как Amazon Alexa и Google Assistant, становятся особенно популярными в преподавании языков. Они предоставляют возможность задавать вопросы по грамматике, лексике или культурным аспектам целевого языка и получать ответы в любое время. Это позволяет практиковать и совершенствовать языковые навыки в реальной и интерактивной среде. Например, вы слушаете голосового помощника, имитируя звучание и интонацию слов и предложений. Использование голосовых помощников позволяет получать немедленную и конструктивную обратную связь.

За последние годы такие инструменты перевода, как Google Translate и DeepL, достигли огромного прогресса благодаря использованию искусственного интеллекта. Эти инструменты помогут вам лучше понимать тексты на иностранном языке и писать собственные тексты. Однако, их следует рассматривать как вспомогательное, а не единственное средство обучения [4].

В заключение, можно утверждать, что искусственный интеллект, как и цифровые инструменты обладают значительным потенциалом для развития и способны адаптироваться к индивидуальным темпам обучения. Следует быть открытым к интеграции искусственного интеллекта в преподавании языка и задуматься о том, насколько эффективным и ценным может быть обучение с его поддержкой в качестве дополнения к традиционным методам обучения.

ЛИТЕРАТУРА

1. Digitales Lernen: Methoden, Vorteile und Nachteile digitaler Bildung in der Schule? [Электронный ресурс] / KITA.DE. – URL: <https://www.kita.de/wissen/digitales-lernen/> (дата обращения: 28.06.2025).
2. Künstliche Intelligenz: Was sind die Vorteile und Nachteile von KI? [Электронный ресурс] / Gründer.de. – URL: <https://clck.ru/3Mm9Xt> (дата обращения: 10.06.2025).
3. Soziale Medien und digitale Lernmöglichkeiten im Fremdsprachenunterricht [Электронный ресурс] / GRIN. – URL: <https://clck.ru/3N6UDJ> (дата обращения: 18.06.2025).
4. Was ist künstliche Intelligenz? [Электронный ресурс] / Medium. – URL: <https://clck.ru/3Mm9nu> (дата обращения: 02.07.2025).

УДК 81

Сироткина Татьяна Александровна
доктор филологических наук,
профессор кафедры филологического
образования и журналистики
Сургутский государственный
педагогический университет
г. Сургут, Россия

Tatiana Sirotkina
Doctor of Philology, Professor of the
Department of Philological Education
and Journalism
Surgut State Pedagogical University
Surgut, Russia
sirotkina71@mail.ru

КОМПАРАТИВНЫЙ АНАЛИЗ ТЕКСТОВ НА ЗАНЯТИЯХ ПО РУССКОМУ ЯЗЫКУ КАК ИНОСТРАННОМУ

В статье рассматривается возможность использования компаративного метода в учебном пособии по русскому языку как иностранному на примере пособия «Россия и Словакия: диалог культур», подготовленного к печати преподавателями университетов Банской Бистрицы и Сургута. Данное учебное пособие содержит тексты для чтения и задания для словацких студентов, изучающих русский язык. Его целью является не только помощь студентам в овладении разными видами речевой деятельности (говорение, слушание, письмо, чтение), но и погружение их культуру двух стран – России и Словакии. Рассматриваются примеры заданий, которые предлагаются студентам при изучении тем «Архитектура», «Живопись», «Кино», «Кухня», «Литература», «Музеи», «Музыка», «Праздники» «Фольклор». На основе анализа материалов пособия делается вывод о том, что именно на основе компаративного анализа текстов обучающиеся смогут не только познакомиться со спецификой культуры страны изучаемого языка, но и воспринять ее в контексте знаний о своей культуре.

Ключевые слова: компаративный анализ; русский язык как иностранный; диалог культур; виды речевой деятельности; учебное пособие.

COMPARATIVE ANALYSIS OF TEXTS IN CLASSES ON RUSSIAN AS A FOREIGN LANGUAGE

The article considers the possibility of using the comparative method in a textbook on Russian as a foreign language using the example of the textbook "Russia and Slovakia: Dialogue of Cultures", prepared for publication by teachers of the universities of Banska Bystrica and Surgut. This textbook contains texts for reading and assignments for Slovak students studying Russian. Its purpose is not only to help students master different types of speech activity (speaking, listening, writing, reading), but also to immerse them in the culture of two countries - Russia and Slovakia. Examples of assignments that are offered to students when studying the topics "Architecture", "Painting", "Cinema", "Cuisine", "Literature", "Museums", "Music", "Holidays" and "Folklore" are considered. Based on the analysis of the materials of the manual, a conclusion is made that it is on the basis of a comparative analysis of texts that students will be able not only to get acquainted with the specifics of the culture of the country of the studied language, but also to perceive it in the context of knowledge of their culture.

Key words: comparative analysis; Russian as a foreign language; dialogue of cultures; types of speech activity; teaching aid.

О том, какими принципами необходимо руководствоваться при создании учебных пособий по русскому языку как неродному, мы уже писали в

одной из предыдущих статей [4]. Сегодня предлагаем рассмотреть один из важных подходов, который нельзя обойти стороной при создании учебных пособий по русскому языку как иностранному.

Одним из важных подходов в лингводидактике, как известно, является компаративный, позволяющий изучать языки в их сопоставлении [1, с. 81]. Элементы компаративного обучения русскому языку [2, с. 195] и русской литературе [3, с. 601] в сопоставлении с родным языком и литературой используются методистами разных стран. Целью нашей статьи является демонстрация возможностей использования компаративного подхода в учебном пособии по русскому языку как иностранному.

Теоретическим материалом для настоящей статьи послужили работы методистов, занимающихся рассмотрением вопросов межкультурного диалога в учебном процессе. Для анализа практических возможностей использования данного метода использовалось находящееся в печати учебное пособие для иностранных студентов «Россия и Словакия: диалог культур».

Данное учебное пособие содержит тексты для чтения и задания для словацких студентов, изучающих русский язык. Целью издания является не только помощь студентам в овладении разными видами речевой деятельности (говорение, слушание, письмо, чтение), но и погружение их культуру двух стран – России и Словакии.

Рассмотрим, какие культуроведческие темы предлагается рассмотреть обучающимся.

Тема 1 – «Архитектура».

Выполняя первое задание данной темы, студенты читают тексты о двух церквях – церкви Преображения Господня в Кижах и церкви Архангела Михаила в Ладомировой и пытаются ответить на вопрос, что сближает эти два архитектурных памятника.

Второе задание связано с анализом текстов о мостах – стихотворения Владимира Паевского о Дворцовом мосте в Санкт-Петербурге и отзыва туристов о братиславском мосте SNP. Обучающиеся описывают, используя прилагательные, сами эти мосты, а также эмоции авторов данных текстов.

Следующий объект – мемориалы. Студенты слушают песню «На Мамаевом кургане тишина», вставляя в текст пропущенные слова, и читают текст о мемориале «Славин» в Братиславе, вспоминают, какие еще национальные памятники культуры Словакии они знают.

Еще один архитектурный объект – дворцы. Обучающимся дается задание прочитать интересные факты о Зимнем дворце Петербурга, объясняя значение выделенных слов, а также послушать текст о Дворце примаса в Братиславе и ответить на вопросы: «Кто спроектировал дворец? Какой стиль лег в основу создания дворца? Какого цвета фасад дворца?».

Следующее задание данной темы связано с чтением текстов о воротах – Петровских воротах Петропавловской крепости в Санкт-Петербурге и Воротах Сигизмунда, самых старых воротах замка Братиславы. Читая данные тексты, студенты знакомятся с семантикой таких архитектурных терминов, как фронтон, барельеф, готический стиль и др.

Основой для выполнения шестого задания темы «Архитектура» послужили тексты о замках – стихотворение Юрия Нестерова о Выборгском замке и текст о городском замке Банска-Бистрицы.

Наконец, последнее задание данной темы – анализ текстов о соборах, Соборе Покрова Святой Богородицы (храме Василия Блаженного) и Соборе святого Мартина в Братиславе. Рассматривая изображение последнего, обучающиеся должны показать, в какой части собора находятся: главный неф, ребристый свод, алтарь, часовни и т.д.

Тема 2 – «Живопись».

Первое задание темы «Живопись» связано с иконописью. Обучающиеся вспоминают, что такое икона, и рассматривают один из ключевых иконописных сюжетов – изображение Святой Троицы.

Следующее задание связано с углублением представлений о портретной живописи. На примере картин А. П. Антропова «Портрет Петра I» и Петера Богуня «Портрет Яна Францисци в роли капитана словацких добровольцев» студенты пытаются найти отличия между портретами, написанными в разных стилях.

Материалом для выполнения третьего задания послужили пейзажные работы Сильвестра Щедрина «Вид с Петровского острова в Петербурге» и Войтеха Климковича «На Страстях Господних», «Развалины Чардаша». Обучающиеся не просто описывают данные картины, но и пытаются сформулировать, какие эмоции они вызывают.

Следующее задание – анализ картин художников-реалистов. Обучающиеся пишут сочинение-описание по картине Василия Перова «Тройка» или картине Доминика Скутецкого «Рынок в Банска-Бистрице».

Последнее задание темы связано с рассмотрением картин модернистов. Студенты читают текст о картине Михаила Врубеля «Царевна-лебедь» и картине Микулаша Голанды «Мать», пытаются ответить на вопрос, какими средствами художники передают на своих полотнах настроение, эмоции.

Тема 3 – «Кино».

Первое задание данной темы основано на анализе текстов о фильмах «Кавказский пленник» и «Магазин на главной улице», после чтения которых обучающиеся отвечают на вопросы.

Второе задание – анализ текстов, объединенных темой «Кино в моей жизни». Студенты читают несколько эссе и пишут свое, представляя в нем личное отношение или к любимому фильму, или к кино вообще.

Выполнение третьего задания начинается с анализа текста о сказке «Золушка», вышедшей на экраны в 1947 г. Обучающиеся знакомятся с сюжетом и некоторыми эпизодами фильма, отвечают на вопрос о том, почему эта сказка оказалась особенно важной для людей в послевоенное время. Вторая часть задания – чтение и анализ текста о фильме «Три орешка для Золушки». После чтения обоих текстов студенты работают с кадрами из фильмов – определяют, какой кадр к какому фильму относится.

Тема 4 – «Кухня».

Первое задание предлагает обучающимся поработать с текстами о русском борще (он представлен в виде инфографики) и словацкой капустнице, рецепт которой на русском языке предлагается написать студентам.

Второе задание связано с анализом текстов о русских блинах, пословицы о которых попутно предлагает им преподаватель, и словацком локше.

Третье задание – чтение текста детской пальчиковой игры «Пельмени», отражающего все этапы приготовления данного блюда. Вторая часть данного задания предлагает словацким студентам написать письмо студентам из России, в котором рассказать о том, как приготовить словацкие клецки.

Еще одно задание данной темы – просмотр видео о приготовлении печенья «Хворост» и рассказ на русском языке о традициях изготовления словацкого печенья «Медвежьи лапки».

Тема 5 – «Литература».

Первое задание темы связано с творчеством двух поэтов – А. С. Пушкина и П. О. Гвездослава, которого называют «словацким Пушкиным».

Второе задание – анализ текстов о творчестве В. В. Маяковского. Обучающимся предлагается познакомиться с биографией поэта, прочитать его стихотворение «Тучкины штучки», а также отрывок из статьи словацкого литературоведа Антона Репоня, посвященной творчеству поэта в чехословацком научном и литературном контексте.

Тема 6 – «Музеи».

В данном разделе пособия представлено несколько заданий, связанных с анализом текстов о российских и словацких музеях. Одно из них, например, предполагает чтение текстов о музее-заповеднике А. С. Пушкина «Михайловское» и литературном музее имени А. С. Пушкина в Бродзянах. Другое задание – анализ текстов о русском Эрмитаже и Словацкой национальной галерее. Студенты вспоминают, картины каких словацких художников демонстрируются в залах галереи, и пишут небольшой текст на русском языке об одном из них.

Тема 7 – «Музыка».

Первое задание темы связано с творчеством российских бардов. Обучающиеся объясняют значение слова «бард», слушают песню Булата Окуджавы «Ах Арбат, мой Арбат» и пытаются доказать или опровергнуть актуальность ее темы. Вторая часть задания – анализ текстов о словацком музыканте Робо Григорове. Студенты слушают его песню «Маяковский» и формулируют ответ на вопрос, почему она стала хитом 90-х годов прошлого века.

Второе задание связано с чтением короткой биографии П. И. Чайковского, а также текста о творчестве Эугена Сухоня и его опере «Водоворот» («Крутнява»). После знакомства с данными текстами студенты отвечают на вопросы о том, какое место занимает классическая музыка в их жизни, каких представителей классической музыки в Словакии они знают.

Таким образом, одним из ведущих принципов при создании учебного пособия по русскому языку как иностранному должен являться компаративный принцип. Именно на основе межкультурного диалога обучающиеся смогут не только познакомиться со спецификой культуры страны изучаемого языка, но и воспринять ее в контексте знаний о своей культуре.

ЛИТЕРАТУРА

1. Сироткина Т. А. Принципы создания учебно-методического пособия по русскому языку как неродному // Вестник Сургутского государственного педагогического университета. 2024. № 1. С. 62–66.
2. Бокова Т. Н., Цыбанева В. А. Компаративный подход в лингводидактике: сущность и перспективы // Образование и саморазвитие. Том 19. 2024. № 1. С. 81–85.
3. Савченко И. А., Алаен Ж. Л., Барсегян С. С. Русский язык в современной Армении: элементы компаративного обучения // Вестник Толмского государственного университета. 2022. № 483. С. 195–199.
4. Садыкова П. Р. Технология развития межкультурной коммуникативной компетенции посредством компаративного анализа зарубежной и русской литературы // Молодой ученый. 2019. № 23. С. 601–605.

УДК 37.017

Тройникова Екатерина Валентиновна

доктор педагогических наук, профессор
кафедры перевода и прикладной
лингвистики (английский и немецкий
языки)

Удмуртский государственный
университет
г. Ижевск, Россия

Ekaterina Troinikova

D. Sc. (Pedagogy), Professor
at the Department of Translation
and Applied Linguistics (English
and German)

Udmurt State University
Izhevsk, Russia
troinikovae@yandex.ru

ОСОБЕННОСТИ ЦИФРОВОЙ ТРАНСФОРМАЦИИ МЕЖКУЛЬТУРНОГО ЛИНГВООБРАЗОВАНИЯ

В статье рассматриваются особенности цифровизации межкультурного лингвообразования. Исследовательское внимание обращено на изучение информационных аспектов данного процесса. Описаны общие и специфические особенности «информатизации» процесса подготовки студентов к межкультурному общению. К общим особенностям информатизации межкультурного языкового образования отнесены социокультурные, антропосообразные (когнитивные, коммуникативные и психоэмоциональные), лингвистические (изменение языковых форм межкультурного общения) особенности. Специфические особенности рассмотрены в связи с влиянием информационной среды на культурные пространства (гибридизация культурной идентичности) и процесс межкультурного взаимодействия (достижение межкультурного диалога в цифровой среде в ходе актуализации медиативной деятельности субъектов межкультурного общения). Определены некоторые трудности цифровизации межкультурного лингвообразования и возможности его развития в условиях цифрового поликультурного многоязычного пространства.

Ключевые слова: межкультурное лингвообразование; межкультурная подготовка студентов; цифровизация межкультурного языкового образования; социокультурные пространства в цифровой среде; цифровая трансформация языкового образования.

FEATURES OF DIGITAL TRANSFORMATION OF INTERCULTURAL LINGUISTIC EDUCATION

The article examines the features of digitalization of intercultural language education. The research focus is on studying the information aspects of this process. General and specific features of "informatization" of the process of preparing students for intercultural communication are described. The general features of informatization of intercultural language education include socio-cultural, anthropological (cognitive, communicative and psycho-emotional), linguistic (change in linguistic forms of intercultural communication) features. Specific features are considered in connection with the influence of the information environment on cultural spaces (hybridization of cultural identity) and the process of intercultural interaction (achieving intercultural dialogue in the digital environment in the course of actualizing the mediation activities of subjects of intercultural communication). Some difficulties in the digitalization of intercultural linguistic education and the possibilities of its development in the context of a digital multicultural multilingual space are identified.

Key words: intercultural linguistic education; intercultural training of students; digitalization of intercultural language education; socio-cultural spaces in the digital environment; digital transformation of language education.

Цифровизация межкультурного лингвообразования обеспечивается созданием цифровой поликультурной многоязычной образовательной среды и включением цифровых технологий в процесс подготовки обучающихся к межкультурному общению. С одной стороны, рассматриваемый процесс существенным образом улучшает возможности получения культуросообразной информации, расширяет формы участия студентов в межкультурном взаимодействии, усиливает образовательный потенциал традиционных межкультурных образовательных технологий за счет их цифровой модификации, способствует формированию «мягких» навыков (например, критического и системного мышления, цифровой грамотности, коллаборативным способам взаимодействия и пр.). С другой стороны, следует отметить, что цифровизация образования вообще и межкультурного лингвообразования в частности, оказывает многоаспектное воздействие на различные виды развития студентов: интеллектуальное, нравственное, физическое, психоэмоциональное и др., формируя ряд социокультурных, психологических и социально-психологических и других угроз-рисков [1].

Отметим, что несмотря на комплексный характер цифровизации, исследователи межкультурной лингводидактики, как правило, ограничивают свою деятельность на разработке вопросов использования цифровых, информационно-коммуникационных технологий в межкультурной языковой подготовке студентов [2], которая (подготовка) ориентирована преимущественно на обучение межкультурной коммуникации как процессу взаимного обмена информацией. В то же время, при изучении данного вопроса следует обратить внимание, что синонимом к термину «цифровизация» является понятие «информатизация», в котором выделяется собственно технологический и информационный аспекты. Последний связан с процессом поиска, обработки, хранения и передачи информации. В этом случае основной риск цифровизации образования вообще и межкультурного лингвообразования в частности состоит в преобладании технологического аспекта в данном процессе безотносительно рассмотрения вопросов подготовки студентов к комплексному процессу обработки информации. Именно в своем взаимодополняющем целостном рассмотрении технологический и информационный аспекты способны обеспечить цифровизацию различных типов образования, гарантировать эффективность жизнедеятельности личности в цифровом поликультурном многоязычном мире. В рамках данной статьи планируется рассмотреть особенности цифровизации межкультурного лингвообразования с учетом собственно информационного аспекта данного процесса, которые (особенности) условно разделены на общие и специфические следующим образом.

Общие особенности цифровизации межкультурного лингвообразования рассмотрены относительно факторов влияния цифровой среды на общество, человека и образование, что представляет определенные возможности развития данного типа образования.

Социокультурные особенности. Цифровизация общества – это «современная форма глобального процесса социальной модернизации, которая дает российскому обществу совокупность новых инструментов и технологий, обеспечивающих возможность масштабного применения нового типа социальной коммуникации» [3, с. 21]. Многочисленные исследования цифровизации современного общества информации (С. А. Дементьев, В. И. Грачев, К. К. Колин и др.) показывают, что цифровая среда значительным образом влияет на социокультурную безопасность общества. Социокультурная безопасность в цифровой среде понимается нами как интегративное единство различных видов безопасностей (аксиологической, информационной, экономической, технологической, межкультурной, социо-психологической, лингвистической и пр.), что предполагает создание и реализацию системы мер, методов и средств правового, организационного, технологического, научно-образовательного характера, обеспечивающих предотвращение негативного влияния информационной среды на гражданское общество и его представителей, ускорение экономического роста государства, регулирование гармоничных, неконфликтных, прогрессивных общественных отношений как внутри государств, так и между ними. В данном интегративном единстве социокультурной безопасности активно изучается аксиологическая безопасность общества и ее отдельных граждан. К ключевым проблемам ее обеспечения относят: смешение, подмену, дискредитацию, нивелирование общегосударственных и национальных ценностей, искажение структуры культурно-исторических событий, декомпозицию культурно-исторического опыта и способов его передачи молодому поколению. Основным средством влияния на аксиосферу на государственном уровне выступают многообразные информационно-технологические и информационно-психологические средства воздействия. Отсюда возникает интерес к способам реализации в межкультурном лингвообразовании одного из базовых принципов обеспечения информационной безопасности общества, среди которых выделим принцип информационной ответственности. Данный принцип основан «на базовых положениях морали и права, действующих в конкретном обществе с точки зрения его социальной зрелости и социокультурных особенностей» [4, с. 81]. Информационная ответственность в межкультурном общении предполагает овладение студентами определенными приемами обеспечения достоверности культуросообразной информации и соблюдения этических норм ее репрезентации согласно различным национально-культурным стандартам, учета «бинарных» (то есть в проекции родной и взаимодействующих культур) особенностей размещения публикаций и комментариев в информационных сетях, обеспечения паритетного взаимодействия культурно-языковых сообществ путем объяснения аксиологических доминант в бинарной перспективе.

Антропосообразные особенности. Обобщение ряда исследований по вопросам влияния цифровой среды на человека [5; 6] показало, что данное влияние носит разновекторный антропотрансформационный характер. В кратком виде такие трансформации можно представить следующим образом:

а) изменения когнитивной сферы и опыта познавательной деятельности: модификация способов восприятия, обработки и перекодировки информации; необходимость мгновенной ориентации в полифонии форм репрезентации информации; перенос акцента с понимания формы и значения слова к пониманию смысла системы информационного блока; нелинейность и фрагментарность информации, скорость ее изменения создает эффект информационного шума, приводит к формированию «клипового мышления» и одновременно к «многозадачности» осуществления познавательного действия, что вызывает быструю утомляемость и истощаемость когнитивных ресурсов личности; и пр.);

б) изменения коммуникативного опыта в ходе освоения новых цифровых форматов общения, в том числе в информационной поликультурной многоязычной среде: появляются новые формы сетевого общения, новые формы индивидуального коммуникативного самовыражения и самопрезентации человека, появляется необходимость постоянного учета национально-культурной локализации сайтов, блогов, влогов и пр.; приоритет нарративности в коммуникации, что выражается в письменно-устном общении и преобладании информационных и оценочных речевых жанров (описание, комментирование, (само)оценка и пр.);

в) изменения в психоэмоциональной сфере связаны с описанными ранее когнитивными и коммуникативными трансформациями, а именно: ограничение реального межличностного общения, изменение его форм в цифровой среде опосредованно влияет на страх установления реального дружеского общения, кратковременность и сетевой (горизонтальный, поверхностный, индивидуализированный) характер контактов. В силу влияния информационно-игровой среды проявляются различные формы аддиктивного поведения, неумение преодолевать коммуникативные трудности в реальном общении при его (общении) нацеленности на диалог личностных позиций, слабая выраженность адекватной рефлексивной самооценки и оценки партнера по общению в информационной среде. Кроме того, происходит снижение порога психологической внушаемости и зависимости, размывание личностной идентичности, психологическое неблагополучие, деструктивное коллективное поведение [1].

Представленные антропосообразные трансформации не исчерпывают себя описанными выше, более того, они постоянно дополняются новыми за счет появления новых геймофицированных и цифровых сред, технологий, помощников, роботов, нейросетей и пр. При этом выделим, что для межкультурного лингвообразования первостепенную роль играет разработка технологий, направленных на овладение студентами комплексом информационных

стратегий, позволяющих противостоять хаотичному, неосознанному, безграничному информационному пространству и обеспечить информационную безопасность на индивидуальном уровне.

Лингвистические особенности. Цифровые дискурсивные формы, новые языковые форматы общения в цифровой среде ведут к значительному изменению лексических и грамматических конструкций, появлению новых стилистических форм презентации информации, определяющих формат общения (формальный, неформальный) и в некоторых случаях затрагивающих процессы понимания/непонимания информации, размещенной в цифровой среде. Трансформация функционирования языка в данной среде связана с его упрощением, распространением интернет-сленга, нецензурной и деструктивной речи. Языковые формы выступают не только как медиаторы выражения собственно смысла информации, но и как инструменты манипулирования сознанием человека, его дезинформации, цифровой агрессии и национального радикализма. Обратим внимание на тот факт, что доминирующую лингвистическую позицию в цифровой среде занимает английский язык, что также нарушает лингвистическую безопасность отдельных государств, в том числе и России. В этом случае в подготовке студентов к межкультурному общению необходимо учитывать лингвистический аспект цифровой коммуникации, что подразумевает овладение обучающимися стратегиями анализа родного и иноязычного интернет-дискурса, форм языковой и семиотической репрезентации культурно-языковых сообществ в цифровой среде (язык и семиотика «национальных» сайтов различной направленности и проблемы их локализации в родной культуре, язык и семиотика формального и неформального общения в различных культурах в чатах, сетевых облачных ресурсах, мессенджерах, социальных сетях, форумах, онлайн-обсуждениях, блогах, влогах и пр.).

Специфические особенности цифровизации межкультурного лингвообразования, выделение которых произошло на основе изучения факторов влияния цифровой среды на национальные культуры и межкультурное взаимодействие.

Культурные пространства в цифровой среде. Процессы диффузии культурных пространств в цифровой среде приводят к тому, что наряду с традиционными национальными и этническими идентичностями формируется транснациональная локация – пространство гибридной культуры, возникающее вследствие поликультурности и многоязычности современного общества, процессов миграции и формирования национальных диаспор внутри отдельных государств. Это непосредственным образом влияет на представителей культурно-языковых сообществ. В частности, происходит ослабление воздействия исторически сложившегося национально-специфического культурно-языкового опыта на культурное самопознание и самоопределение отдельного индивида. Все чаще культурная идентичность современного человека представляет собой гибридизированный персонифицированный конструкт, созданный на основе взаимодействия нескольких

культурных традиций (см. подробно [7]). В этом случае подготовка студентов к межкультурному общению представляет собой динамическое взаимодействие процессов культурного самоопределения и межкультурной социализации, что реализуется в том числе за счет содержания культуросообразного учебного материала, отражающего специфику национально-культурной и общецивилизационной картины мира и представленного в различных цифровых ресурсах.

Межкультурное взаимодействие в цифровой среде. Цель межкультурного взаимодействия в цифровой среде – расширение доступа к культуросообразной информации, развитие культурного обмена, популяризация культурных традиций, решение глобальных общекультурных цивилизационных проблем. Продуктивной формой данного вида взаимодействия в цифровой среде является межкультурный диалог, который обеспечивается равноправностью и равноценностью каждого лингвоэтносоциума в процессе общения в информационном пространстве. Диалогизация межкультурного взаимодействия обеспечивается медиативной деятельностью субъектов межкультурного общения, что подразумевает информационное моделирование межкультурного диалога. Данный процесс предполагает поиск сферы допустимого согласования позиций взаимодействующих сторон, определение диалоговых смыслообразований и языковых средств их объяснения. В этих условиях межкультурное лингвообразование нацелено на формирование субъекта межкультурного диалога в цифровой среде, готового и способного осуществлять медиативную деятельность на концептуальном, языковом и коммуникативном уровнях. Основу такой деятельности составляет поиск и обработка полимодальной культуросообразной информации, диалоговое смыслообразование и диалоговый нарратив.

Таким образом, описанные особенности цифровизации межкультурного лингвообразования формируют ряд трудностей и новых направлений организации данного процесса. Основные трудности цифровизации межкультурного языкового образования на научно-методическом уровне состоят в некоторой терминологической однозначности понимания данного процесса. В частности, употребление термина «цифровизация» обуславливает доминирование исследований, направленных на изучение процесса интеграции цифровых технологий в процесс межкультурной подготовки студентов. При этом, как было указано выше, цифровизация сопровождается многоаспектным воздействием информационной среды на общество, культуру, образование и собственно человека, формируя таким образом новые задачи в межкультурном лингвообразовании. К таким задачам относятся: изучение возможности использования информационного подхода в организации процесса межкультурной подготовки студентов; интеграция в систему целевых ориентиров подготовки студентов к межкультурному общению в цифровой среде, наряду с информационной грамотностью, группы инструментальных стратегий поиска, организации, бинарной интерпретации и оценки культуросообразной информации, стратегий информационного моделирования и ме-

диации межкультурного диалога в цифровой среде; разработка эффективных методов, приемов и форм создания и/или отбора культуросообразного цифрового учебного материала; освоение и внедрение информационных технологий, позволяющих (само)развивать когнитивный, нравственный, коммуникативный, творческий потенциал личности обучающегося как субъекта межкультурного диалога в цифровой среде.

ЛИТЕРАТУРА

1. Кисляков П. А., Сорокоумова С. Н., Егорова П. А. Психолингвистическая безопасность личности студента и ее обеспечение в процессе обучения в вузе // Язык и культура. 2020. №. 49. С. 199–218.
2. Сысоев П. В., Евстегнеев М. Н. Методика обучения иностранным языкам с использованием новых информационно-коммуникационных Интернет-технологий. Ростов н/Д : Феникс; М. : Глосса-Пресс, 2010. 182 с.
3. Кожевников С. Б., Ермоленко Г. А. Цифровизация современной культуры и гуманитарная рефлексия // Цифровая гуманитаристика: человек в «прозрачном» обществе. 2021. С. 21–28.
4. Дементьев С. А. Принципы как основополагающие постулаты концепции информационной безопасности человека // Философия права. 2019. №. 1 (88). С. 80–84.
5. Белинская Е. П. Информационная социализация в поликультурном пространстве // Изв. Саратов. ун-та. Нов. сер. Сер. Акмеология образования. Психология развития. 2014. Т. 3, вып. 3 (11). С. 253–258.
6. Марцинковская Т. Д. Информационное пространство как фактор социализации современных подростков // Мир психологии. 2010. №. 3. С. 90–102.
7. Тарева Е. Г., Тройникова Е. В. Культурная идентичность субъекта: векторы информационной трансформации // Язык и культура. 2023. № 64. С. 45–64.

УДК 372.881.1

Шпаковская Ирина Николаевна

старший преподаватель кафедры
английского языка гуманитарных
факультетов
Белорусский государственный
университет
г. Минск, Беларусь

Iryna Shpakouskaya

Senior lecturer at the Department
of English for Humanities
Belarusian State University
Minsk, Belarus
shpakouskayain@bsu.by

ЦИФРОВЫЕ ТЕХНОЛОГИИ КАК СРЕДСТВО ОПТИМИЗАЦИИ ПРОЦЕССА ОВЛАДЕНИЯ ГРАММАТИЧЕСКИМ МАТЕРИАЛОМ СТУДЕНТАМИ НЕЯЗЫКОВЫХ ВУЗОВ

Рассматривается использование цифровых технологий в процессе обучения иностранному языку, в частности, на примере применения метода «Перевернутый класс» при изучении грамматического материала студентами юридического факультета Белорусского государственного университета. Раскрываются детали начальных стадий внедрения элементов смешанного обучения с помощью компьютерных средств при овладении грамматическим аспектом языка. Автор делится опытом успешного интегрирования данных технологий в процесс преподавания грамматической темы «Модальные глаголы». Обсуждаются особенности распределения основных этапов (ознакомления, тренировки и речевой практики) обучения данному грамматическому явлению между онлайн-формой преподавания и аудиторными занятиями, также оценивается успешность данного процесса, отмечаются его потенциальные сложности. Поскольку цифровые технологии позволяют сократить время, которое приходилось тратить в аудитории как на объяснение нового материала, так и на отработку навыков владения данным материалом, а также, частично, и на контроль владения материалом, делается вывод о значительном потенциале привлечения информационных технологий при проведении занятий, что имеет особую важность в условиях сокращения количества аудиторных часов по иностранному языку в неязыковых вузах.

Ключевые слова: цифровые технологии; смешанное обучение; «перевернутый класс»; грамматический материал; онлайн коммуникация; очная форма.

DIGITAL TECHNOLOGIES AS TOOLS TO OPTIMIZE THE PROCESS OF MASTERING GRAMMAR BY STUDENTS OF NON-LINGUISTIC UNIVERSITIES

The article deals with the issue of using digital technologies namely within the method of “Flipped classroom” while teaching grammar at the law faculty of BSU. It reveals the details of the first stages of introducing blended learning elements into the process of mastering grammar with the help of computer technologies. The author presents examples of successful integration of the information technologies into teaching grammar topic “Modal verbs” and discusses the peculiarities of distributing main teaching stages (presentation, practice and production) between online and face-to-face classes. The author evaluates the success rate of the given process and points out the potential challenges. As digital technologies contribute to reducing the classroom time otherwise spent introducing the new material, practicing and checking the students’ grasp of

the material, the author arrives at the conclusion that information technologies possess significant potential regarding the process of teaching foreign languages in the situation characterized by reducing classroom hours of English in non-linguistic universities.

Key words: digital technologies; blended learning; flipped classroom; grammar material; online learning; face-to-face learning.

Овладение особенностями системы английского языка в его фонетическом, лексическом и грамматическом аспектах является одной из задач освоения учебной дисциплины «Иностранный язык» в неязыковом вузе.

В последние годы обучение грамматике при преподавании иностранного языка на юридическом факультете БГУ стало своего рода вызовом для преподавателей в связи с сокращением количества аудиторных занятий по предмету. Это в первую очередь затронуло изучение иностранного языка студентами специальности «Политология», поскольку данные студенты изначально не сдают английский язык при поступлении и, как следствие, их уровень владения языком, в частности, его грамматическим аспектом, существенно ниже, чем у студентов других специальностей.

В данных обстоятельствах особую актуальность приобретает использование цифровых образовательных технологий причем и для преподавания английского языка в целом, и таких его аспектов, как грамматика. Цифровые образовательные технологии определяются в литературе как инновационный способ организации учебного процесса, основанный на использовании электронных систем, обеспечивающих наглядность [1, с. 54]

Впервые возможность и необходимость системного, а не спорадического использования цифровых технологий при обучении иностранному языку студентами юридического факультета БГУ появилась у преподавателей кафедры английского языка гуманитарных факультетов более 10 лет назад. В период с 2016 по 2021 г. студентам первого курса юридического факультета в течение первого семестра предоставлялась возможность посещать дополнительные занятия по английскому языку, так называемые «группы выравнивания», основной задачей которых было помочь студентам, испытывающим сложности с языковой подготовкой, ликвидировать «пробелы» в знаниях и прежде всего в плане грамматики. Поскольку данные занятия не являлись частью основной программы и не были внесены в основную сетку расписания, приходилось частично проводить их в формате онлайн еще до того, как такая форма проведения занятий стала повсеместно распространенной из-за пандемии. На тот момент платформа BigBlueButton, рекомендованная к использованию в БГУ, находилась на начальных стадиях использования, нередко возникали сложности с форматом видео-конференции (в том числе из-за проблем с Интернет-связью у студентов), поэтому преимущественно приходилось ограничиваться общением через чат.

В связи с этим преподавателями кафедры английского языка гуманитарных факультетов ФСК БГУ на платформе Moodle был разработан онлайн-курс «Support English Course» («Иностранный язык: выравнивающий факультатив»), который включал в себя большое количество тренировочных

упражнений по грамматическим темам первого курса, а также ряд компьютерных тестов для контроля овладением студентами материалом [2]. Соответственно, при возникновении технических сложностей во время работы видео-конференции, которые на первоначальном этапе имели место на регулярной основе, преподаватели давали задание студентам через чат и следили за его выполнением в режиме онлайн.

К сожалению, в 2022 г. данные группы были упразднены, что значительно осложнило процесс обучения студентов с изначально невысоким базовым уровнем языковой подготовки. Фактически одновременно с этим количество аудиторных часов на юридическом факультете уменьшилось на порядок: сначала были значительно сокращены часы на втором курсе факультета (со 104 аудиторных часов до 82), а затем дисциплина «Иностранный язык» была полностью исключена из преподавания на втором курсе. В настоящий момент английский язык преподается лишь на первом курсе юридического факультета в объеме 152 аудиторных часов. У студентов специальности «Политология» данная цифра еще ниже – всего лишь 136 аудиторных часов.

Отдельно следует упомянуть, что грамматический модуль курса «Иностранный язык (профессиональная лексика)», преподаваемый на специальности «Политология», включает в себя такие темы, как «Система времен», «Страдательный залог», «Косвенная речь», «Модальные глаголы», «Сослагательное наклонение» и «Неличные формы глагола», т. е. очень широкий спектр грамматических явлений, для изучения которых нужно большое количество времени.

Все вышеуказанные темы также являются частью школьной программы, однако, как показывает диагностическое тестирование, проводимое на первом занятии по английскому языку, у подавляющего большинства студентов отсутствуют систематические знания по данным темам. В лучшем случае студенты способны распознать базовые временные формы (активный и пассивный залог) и основные модальные глаголы (такие как *must*, *can*, *may*), но применение данных форм в предложенном контексте, не говоря уже о спонтанной речи, вызывает у подавляющего большинства из них значительные затруднения.

На сегодняшний день, благодаря цифровизации образовательного процесса, одной из самых распространенных и эффективных моделей преподавания информации стала технология *Flipped classroom* 'Перевернутый класс'. Одним из преимуществ использования данной модели является возрастающая ответственность студента, что ведет к развитию таких личностных характеристик, как активность, самостоятельность и инициативность, способствует формированию метапредметных (самоорганизация, навыки управления временными ресурсами) и коммуникативных (взаимодействие с одноклассниками в ходе совместной работы) навыков [3, с. 190].

В рамках технологии «Перевернутый класс» теоретические аспекты, как правило, изучаются студентами самостоятельно в дистанционном формате, в

то время как выполнение практических заданий осуществляется в классе вместе с преподавателем. При применении такого способа обучения обеспечивается успешная коммуникация внутри аудитории, а преподаватель направляет студентов на самостоятельные исследования и совместную работу. Таким образом, студенты являются активными участниками образовательного процесса и могут делиться полученными знаниями, вести активную учебную деятельность внутри группы, а также воспринимать информацию более концентрированно.

В настоящей статье хотелось бы поделиться опытом применения технологии «Перевернутый класс» на примере преподавания грамматической темы «Модальные глаголы». Изучение данной темы можно условно разделить на два блока: употребление модальных глаголов в ситуациях, относящихся к настоящему и будущему, и употребление модальных глаголов в ситуациях прошедшего времени. Важно отметить, что, как показывает практика, изучение (или повторение) первого из вышеназванных блоков не вызывает у студентов значительных затруднений, в отличие от овладения информацией о функционировании модальных глаголов в контексте прошедшего времени. Особую сложность вызывает осознание того факта, что далеко не все глаголы имеют форму прошедшего времени, а в большинстве случаев временная отнесенность к прошедшему передается через форму инфинитива основного глагола. В этом смысле использование технологии «Перевернутый класс», предполагающее самостоятельное изучение правил студентами, представляло определенный риск.

Итак, в рамках «Перевернутого класса» студенты должны были сами изучить теоретическую информацию о формах и правилах употребления модальных глаголов. Для этого в курсе «English for Lawyers: Basic Course» ('Английский язык для юристов: основной курс') [4] им были предложены ссылки на грамматические справочники, а также схемы и видео, которые они могли использовать при подготовке к занятию. После изучения правил студентам также было предложено выполнить ряд тренировочных упражнений, проверяющих, насколько хорошо они усвоили теоретическую часть: они должны были выполнить тест (из 30 пунктов), состоящий из заданий на множественный выбор, а также упражнений типа *True / False* 'Верно / Неверно'. Естественно, результаты выполнения данных заданий были доступны для контроля и анализа преподавателям в режиме онлайн, что позволяло оценить степень готовности студентов к аудиторной работе по теме.

На самом занятии преподавателем проводился блиц-опрос с целью контроля понимания студентами основных теоретических моментов употребления модальных глаголов. Одновременно студенты также имели возможность уточнить или прояснить те или иные положения, изложенные в домашних видео и схемах. Помимо теоретической информации, ответы студентов на вопросы преподавателя должны были сопровождаться примерами конструкций с модальными глаголами из видео на портале (которые

студенты также должны были просмотреть при подготовке к занятию), при этом видео представляло собой компиляцию из известных поп- и рок-песен, например, *It must have been love (Roxette)* ‘Должно быть, это была любовь’. *I should have told you (Fuel)* ‘Мне следовало тебе сказать’. Как вариант, можно было предложить студентам найти примеры из текстов их любимых исполнителей. Вторая альтернатива данному заданию заключалась в возможности проиллюстрировать примеры употребления модальных глаголов цитатами из англоязычных фильмов. Как правило, студенты с интересом и удовольствием реагировали на такого рода задания, особенно в плане поиска и демонстрации своим одноклассникам цитат из любимых фильмов.

Если говорить о практических заданиях, выполняемых непосредственно в классе, то упор делался на трансформационные упражнения, сформулированные следующим образом: *Paraphrase the following statements using modal verbs. The meaning must stay the same.* ‘Перефразируйте высказывание, модальные глаголы. Смысл высказывания должен сохраняться’.

Безусловно, в большинстве своем упражнения, выполняемые в аудитории, должны быть коммуникативно значимыми. Соответственно, студентам предлагались коммуникативно-ориентированные задания на составление монологов и диалогов, например, *Student M was absent from our last class. Make you assumptions regarding the reasons for them missing the class (using the modal verbs)* ‘Студент М. отсутствовал на прошлом занятии. Выскажите свои предположения о причинах пропуска (используя известные вам модальные конструкции)’.

Формой контроля уровня владения студентами грамматической темой «Модальные глаголы» была письменная контрольная работа, которую студенты писали в аудитории на бумажных носителях. При этом в преддверии написания этой работы студентам предлагалось пройти компьютерное тестирование по данной теме, разработанное в электронном курсе «English Grammar» (‘Английская грамматика’) [5]. Задания данного теста охватывают как теоретические аспекты формирования и употребления модальных конструкций, так и практические упражнения (множественный выбор, верно / неверно и т. д.). В итоге, результаты, продемонстрированные студентами при написании контрольной работы по модальным глаголам в 2023–2024 учебном году (первый год применения данной технологии после сокращения часов аудиторных занятий) оказались относительно высокими (6,5 балла), что оказалось на 0,4 балла выше, чем в предыдущем учебном году. В 2024–2025 учебном году данный показатель незначительно повысился до 6,6 балла.

По опыту можно сделать вывод о том, что потенциал цифровых технологий является практически безграничным в рамках образовательного процесса в целом, и в особенности при обучении грамматике в условиях сокращения аудиторных часов. Помимо значительной экономии времени, использование данных технологий способствует росту самостоятельной активности студентов, они учатся мыслить критически, также реализуется более индивидуальный подход при работе с группой. Обратная связь от

студентов (результаты проводимого каждый год анкетирования) свидетельствует о том, что студенты положительно воспринимают совмещение очных форм обучения и обучения онлайн.

Таким образом, использование цифровых технологий в образовании является неотъемлемым компонентом профессионального образования, дающего возможность развивать конкурентоспособные качества студентов на пути к становлению высококвалифицированных специалистов. На сегодняшний день игнорировать образовательный потенциал цифровых технологий невозможно. В связи с этим центральной задачей преподавателя является использование цифровых технологий в образовательном процессе с учетом индивидуальных способностей обучающегося. В целом цифровизация является средством оптимизации структуры образования, а также инновационным подходом обеспечения образовательного процесса.

ЛИТЕРАТУРА

1. Цифровые технологии в образовательном пространстве / О. И. Ваганова, А. В. Гладков, Е. Ю. Коновалова, И. Р. Воронина // Балтийский гуманитарный журнал. 2020. Т. 9, № 2 (31). С. 53–56.
2. Иностранный язык: выравнивающий факультатив (Support English Course) : электронный курс / авт.-сост.: Т. В. Коваленок [и др.]. URL: <https://eduenglish.bsu.by/course/view.php?id=24> (дата обращения: 27.06.2025).
3. Саутина Е. Д. Применение технологии «Перевернутый класс» при обучении грамматическому материалу обучающихся старших классов // Актуальные проблемы филологии и методики преподавания иностранных языков. 2023. Т. 17, № 2. С. 189–197.
4. Иностранный язык (English for Lawyers: Basic Course) : электронный курс / авт.-сост.: Т. А. Катомахина [и др.]. URL: <https://eduenglish.bsu.by/course/view.php?id=6> (дата обращения: 27.06.2025).
5. Иностранный язык (English Grammar) : электронный курс / авт.-сост.: Т. В. Коваленок [и др.] URL: <https://eduenglish.bsu.by/course/view.php?id=4>. (дата обращения: 27.06.2025).

Секция 4.
МУЛЬТИМОДАЛЬНАЯ КОММУНИКАЦИЯ
В ВИРТУАЛЬНОМ ПРОСТРАНСТВЕ

УДК 811.11-112

Долженкова Ольга Михайловна
старший преподаватель
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Volha Dolzhenkova
Senior lecturer
Belarusian State University
of Foreign Languages
Minsk, Belarus
yogi4smile@yahoo.com

КОРРЕЛЯЦИЯ МУЛЬТИМОДАЛЬНЫХ ЭЛЕМЕНТОВ
И ЛИНГВОКОГНИТИВНЫХ СХЕМ И МАТРИЦ ХРОНОТОПА
В АНГЛОЯЗЫЧНОЙ НОВОСТНОЙ ЭКОНОМИЧЕСКОЙ СТАТЬЕ

В статье представлен анализ мультимодальных элементов англоязычной новостной статьи экономической тематики с точки зрения их корреляции с лингвокогнитивной схемой и матрицей текстового хронотопа. Концептуальный анализ показывает, что мультимодальные элементы прямо либо косвенно отражают лингвокогнитивные схемы хронотопа новостной статьи, участвуя в изобразительной, координирующей и интегрирующей функциях текстового хронотопа. Кроме того, мультимодальные элементы могут заполнять слоты лингвокогнитивной матрицы хронотопа, незаполненные вербально. При этом мультимодальные элементы конкретизируют пространственно-временные параметры текста, усиливая вовлеченность индивидуального реципиента. В этом случае мультимодальные элементы новостного текста участвуют в интегрирующей и прагматической функциях хронотопа. Мультимодальность в англоязычных новостных статьях экономической тематики служит инструментом интеграции обобщенных данных, представленных вербально, с индивидуальным опытом реципиента, расширяя текстовый хронотоп.

Ключевые слова: хронотоп; лингвокогнитивная схема хронотопа; лингвокогнитивная матрица хронотопа; мультимодальный элемент; мультимодальность; англоязычная новостная статья; экономический дискурс.

CORRELATION OF MULTIMODAL ELEMENTS
AND LINGUOCOGNITIVE CHRONOTOPE SCHEMATA AND MATRICES
IN ENGLISH LANGUAGE NEWS ARTICLES ON ECONOMICS

The article analyzes multimodal elements in English language news articles on economics through the lens of their correlation with the linguocognitive schema and matrix of textual chronotope. A conceptual analysis reveals that multimodal elements reflect linguocognitive chronotope schemata of news articles either directly or indirectly, contributing to the depictive, coordinating and integrating functions of textual chronotope. Additionally, multimodal elements may fill gaps in the linguocognitive chronotope matrix, which have been left unaddressed verbally. By making spatiotemporal parameters of the text more specific these elements enhance individual reader engagement. In this context multimodal elements in news articles have a role in the integrating and pragmatic functions of textual chronotope. Multimodality in English

language news articles on economics serves as a tool for integrating general data which is represented verbally with personal experience of the recipient, thereby expanding textual chronotope.

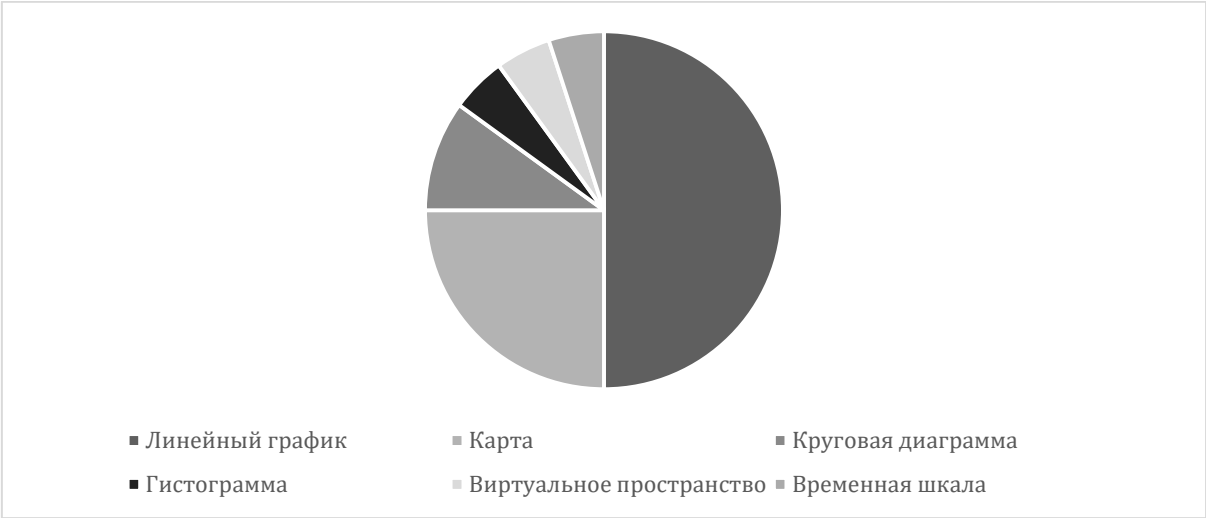
Key words: chronotope; linguocognitive chronotope schema; linguocognitive chronotope matrix; multimodal element; multimodality; English language news article; economic discourse.

Рассмотрение текстового хронотопа как когнитивного инструмента начинается в работах М. М. Бахтина, который наделял художественный хронотоп текстообразующим, сюжетным, изобразительным и ценностным значением и отмечал, что «всякое вступление в сферу смыслов совершается только через ворота хронотопов» [1, с. 193]. Б. Койнен развивает идею М. М. Бахтина о когнитивном потенциале хронотопа и проводит параллели между художественным хронотопом и схемами запоминания Ф. Бартлетта, а также суперструктурами Т. ван Дейка [2]. Подобный подход позволяет расширить область изучения хронотопа и включить в исследование разнообразные нехудожественные тексты. К функциям текстового хронотопа Е. Н. Широкова в традиции М. М. Бахтина относит координирующую (определение пространственно-временной соотнесенности событий, явлений, концепций), интегрирующую (объединение элементов текста путем их соположения в пространстве и времени, формирование виртуальной реальности текста), изобразительную и аксиологическую [3, с. 183].

Задачей настоящей работы является выявить корреляцию лингвокогнитивных схем и матриц хронотопа с мультимодальными элементами англоязычной новостной статьи экономической тематики, а также определить роль мультимодальных элементов в построении хронотопа как ментальной схемы текста. На современном этапе развития коммуникации, по словам Л. И. Гришаевой, «невербальные способы сообщения информации реципиенту становятся все более функционально нагруженными» [4, с. 90]. Мультимодальные элементы новостных экономических статей включают в себя в первую очередь фото, иллюстрации (чаще диаграммы различного типа) и видео. К мультимодальным элементам можно также отнести предоставляемую отдельными новостными сервисами (The Wall Street Journal, Bloomberg) возможность прослушать текст статьи вместо чтения.

Анализ хронотопа 20 новостных статей экономической тематики из онлайн-версий британских и американских новостных сервисов проводился методами дефиниционного и контекстуального анализа, лингвостилистического и интерпретационного анализа текста, концептуального анализа. Определены лингвокогнитивные схемы хронотопа, характерные для англоязычной новостной экономической статьи, построена лингвокогнитивная матрица хронотопа изучаемого типа текста (на основе индивидуальных матриц каждого текста). Учитывались как физические хронотопы, отображающие реальные физические пространственно-временные параметры описываемой ситуации, так и моделирующие, основанные на концептуальных пространственных метафорах и представляющие какую-либо систему метафорически в пространственных терминах.

По результатам анализа, каждый новостной текст имеет одну основную лингвокогнитивную схему хронотопа, может также быть одна или две вспомогательных. Выявлены следующие лингвокогнитивные схемы хронотопа: *линейный график* (основная схема в 10 статьях, 50 %), *физическая, политическая и экономическая карта* (основная схема в 5 статьях, 25 %), *круговая диаграмма* (основная схема в 2 статьях, 10 %), *столбчатая диаграмма* (основная схема в 1 статье, 5 %), *виртуальное пространство* (основная схема в 1 статье, 5 %), *временная шкала* (основная схема в 1 статье, 5 %) (Рисунок).



Распределение основных лингвокогнитивных схем хронотопа
в новостной экономической статье

Сопоставление лингвокогнитивных схем хронотопа с мультимодальными компонентами позволило выявить 2 типа их корреляции: прямое и косвенное отображение схемы хронотопа в мультимодальных компонентах.

Во-первых, мультимодальные компоненты могут прямо отображать основную лингвокогнитивную схему хронотопа. Чаще всего это происходит, когда схема хронотопа представляет собой линейный график, изменение экономических показателей во времени, и сам график представлен в статье (6 статей из 10 со схемой *линейный график*). Например, в статье *Chocolate prices soar as UK inflation stays at highest in over a year* 'Цены на шоколад взлетели, уровень инфляция в Великобритании самый высокий за год' (BBC, 18 июня 2025) основной лингвокогнитивной схемой хронотопа является *линейный график*, показывающий изменение цен и уровня инфляции. Эта схема отражается в заголовке статьи и в приведенном в статье линейном графике изменения цен на шоколад с января 2016 по май 2025.

Лингвокогнитивная схема *карта* репрезентируется как в фотографиях, так и в диаграммах. Например, в статье *Carmakers win reprieve from Canada and Mexico tariffs* ‘Автопроизводители получили отсрочку по канадским и мексиканским пошлинам’ (BBC, 6 марта 2025) схема *физическая, политическая и экономическая карта Северной Америки* представлена фотографией автомобилей на дороге, над которыми развевается канадский флаг (возможно, это отсылка к пересечению границы между Канадой и США). Кроме того, схема хронотопа отображается в видео с подписью *I'm getting angry and anti-American – Canadians on tariff threat* ‘«Я злюсь и настраиваюсь против Америки» – канадцы об угрозе пошлин’. Две страны, находящиеся в Северной Америке упомянуты в подписи к видео, причем подчеркнута их враждебность друг к другу, и та же идея развивается в самом видео, которое представляет негативные комментарии канадцев о планируемом введении торговых пошлин.

Во-вторых, мультимодальные элементы могут косвенно отражать основную либо вспомогательную лингвокогнитивную схему хронотопа. Например, в статье *China's manufacturing activity shrinks as US tariffs take effect* ‘Китайское производство сокращается по мере того, как США вводит пошлины’ (Financial Times, 30 апреля 2025) вспомогательной схемой хронотопа является *карта*, на фото мы видим рабочего в спецодежде, на каске надпись на китайском языке. Т.е., косвенно фото отображает страну, находящуюся в фокусе хронотопа и статьи в целом. Прямого отображения вспомогательной схемы хронотопа мультимодальными элементами в нашей выборке не обнаружено. Сводная таблица репрезентации лингвокогнитивных схем хронотопа в мультимодальных элементах новостных статей представлена ниже (Таблица 1).

Таблица 1

Репрезентация лингвокогнитивных схем хронотопа
в мультимодальных элементах англоязычных новостных статей
экономической тематики

Основная лингвокогнитивная схема – линейный график (10 статей)		
Прямое отображение основной схемы (график/гистограмма) – 6 статей	Вспом. схема – карта (6 статей)	
	Косвенное отображение (фото) – 3 статьи	Не отображена – 3 статьи
Косвенное отображение основной схемы (фото) – 2 статьи	Вспом. схема – круговая диаграмма (2 статьи), не отображена	
Основная схема не отображена – 2 статьи	Нет вспом. схемы – 2 статьи	

Основная лингвокогнитивная схема – карта (5 статей)		
Прямое отображение основной схемы (фото с флагами стран, фото города, диаграммы с указанием стран) – 4 статьи	Вспом. схема – круговая диаграмма (3 статьи)	
	Косвенное отображение (фото) – 2 статьи	Не отображена – 1 статья
	Вспом. схема – линейный график (3 статьи), не отображена	
Косвенное отображение основной схемы (фото) – 1 статья	Нет вспом. схемы – 2 статьи	
Основная лингвокогнитивная схема – круговая диаграмма (2 статьи)		
Косвенное отображение основной схемы (фото) – 2 статьи	Вспом. схемы – карта и линейный график (1 статья), не отображены	
	Нет вспом. схемы – 1 статья	
Основная лингвокогнитивная схема – виртуальная реальность (1 статья)		
Прямое отображение основной схемы (фото)	Вспом. схема – карта, не отображена	
Основная лингвокогнитивная схема – временная шкала (1 статья)		
Косвенное отображение основной схемы (фото)	Вспом. схемы – линейный график и круговая диаграмма, не отображены	
Основная лингвокогнитивная схема – столбчатая диаграмма (1 статья)		
Косвенное отображение основной схемы (фото)	Нет вспом. схемы	

Таким образом, важной хронотопической функцией мультимодальных компонентов в новостной статье является репрезентация лингвокогнитивной схемы хронотопа, при этом мультимодальность участвует в изобразительной, координирующей и интегрирующей функциях хронотопа.

Кроме лингвокогнитивной схемы, выявлена также лингвокогнитивная матрица хронотопа англоязычной новостной экономической статьи. Она включает 5 уровней пространственного измерения: глобальный, континент, страна/регион страны, компания/индивид, виртуальный. На временной оси матрицы выявлены 6 компонентов: историческая справка, непосредственная предыстория, описываемое событие, современное состояние, перспективы, общая панхроничная информация. Репрезентация каждого элемента матрицы в исследуемых статьях представлена в Таблице 2. Первая цифра обозначает общее количество статей, в которых представлен данный слот матрицы, вторая цифра (полужирным шрифтом) обозначает количество статей, в которых этот слот находится в фокусе внимания (отражен в заголовочном комплексе [5, с. 369], сильных позициях статьи, мультимодальных компонентах, имеет большее количество пространственно-временных маркеров по сравнению с другими слотами).

Таблица 2

Лингвокогнитивная матрица хронотопа англоязычной новостной экономической статьи (всего статей, в которых представлен данный слот матрицы/статей, где данный слот матрицы в фокусе).

	Историческая справка	Непосредственная предыстория	Описываемое событие	Современное состояние	Перспективы	Общая панхроничная информация
Глобальный	7	13/7	6/4	12/5	13/5	—
Континент	2/1	4/2	4/1	5/2	6/4	—
Страна/ регион страны	8/1	10/6	12/8	14/7	14/6	5
Компания/ индивид	6/2	9/3	7/1	8/1	6/3	—
Виртуальный	—	1	3	2	2	—

Второй распространенной хронотопической функцией мультимодальных элементов является заполнение одного из слотов матрицы хронотопа, который не заполнен вербальными пространственно-временными маркерами в тексте. Чаще всего это индивидуальный уровень пространственного измерения (7 статей, 35 %) или уровень «историческая справка» временного измерения (2 статьи, 10 %).

Например, в статье *The Surge in Ocean Shipping Rates Is Peaking* «Пост тарифов на океанские перевозки достиг максимума» (The Wall Street Journal, 17 июня 2025) слот «историческая справка» на глобальном уровне схемы хронотопа представлен только графиком, данные на котором начинаются с июня 2024 года, за год до основных временных параметров событий, описанных в тексте. В статье *Americans cut back sharply on their spending last month amid tariffs* ‘Американцы резко сократили расходы в прошлом месяце на фоне пошлин’ (CNN, 17 июня 2025) фото и подпись к нему являются единственным наполнением слота «описываемое событие» на индивидуальном уровне. Лингвокогнитивная схема хронотопа статьи – *линейный график*, отображающий изменение уровня розничных продаж в США. Индивидуальный уровень пространственного измерения хронотопа не представлен в тексте статьи, но отображен на фото с подписью *A customer shops for toys at a big box retailer on May 12 in Chicago* ‘Покупательница выбирает игрушки

в большом розничном магазине 12 мая в Чикаго'. Фото и подпись к нему расширяют текстовый хронотоп до уровня отдельного магазина и индивидуального потребителя, с которым может себя идентифицировать читатель. Таким образом, заполняя слоты индивидуального уровня лингвокогнитивной схемы хронотопа мультимодальные элементы текста выполняют интегрирующую функцию, а также прагматическую функцию привлечения внимания читателя, приближения общеэкономической темы статьи к интересам индивидуального реципиента.

Подводя итог, отметим, что мультимодальные элементы выполняют в англоязычных новостных статьях экономической тематики две важные хронотопические функции. Во-первых, эти элементы, особенно диаграммы, прямо или косвенно отражают основную лингвокогнитивную схему хронотопа, а также могут косвенно отражать вспомогательную схему хронотопа, выполняя изобразительную, координирующую и интегрирующую функции. Во-вторых, мультимодальные элементы, чаще фото, заполняют слоты лингвокогнитивной матрицы хронотопа, незаполненные на вербальном уровне, дополняют текстовый хронотоп индивидуальным уровнем, расширяют его, приближая текст к каждому отдельному читателю, и выполняя интегрирующую функцию, а также прагматическую функцию привлечения внимания реципиента.

ЛИТЕРАТУРА

1. Бахтин М. М. Формы времени и хронотопа в романе // Эпос и роман. 2000. С. 9–194.
2. Keunen B. Bakhtin, Genre Formation, and the Cognitive Turn: Chronotopes as Memory Schemata [Electronic resource] // CLCWeb: Comparative Literature and Culture. 2000. Vol. 2, № 2.
DOI: <https://doi.org/10.7771/1481-4374.1069>
3. Широкова Е. Н. Хронотоп текста, хронотоп дискурса (к вопросу разграничения текстовых и дискурсивных категорий) // Научное наследие Б. Н. Головина в свете актуальных проблем современного языкознания : (к 100-летию со дня рождения Б. Н. Головина): сб. ст. по материалам Междунар. науч. конф., Нижний Новгород, 28-30 сент. 2016 г. Нижний Новгород: Издательство "Деком", 2016. С. 179–184.
4. Гришаева Л. И. Мультимодальность, гибридизация, карнавализация в организации текста и уплотнение информационного потока // Немецкая филология в Санкт-Петербургском государственном университете. 2022. Вып. 12. С. 66–101.
5. Скубач Л. Н. Специфика заголовков в новостных текстах интернет-СМИ // Язык в эпоху цифровых трансформаций и развития искусственного интеллекта : сб. науч. ст. по итогам междунар. науч. конф., Минск, 23–24 окт. 2024 г. / редкол.: Н. Е. Лаптева (отв. ред.) [и др.]. Минск : МГЛУ, 2024. С. 369–373.

УДК 81'34

Евдокимова Вера Вячеславовна

кандидат филологических наук,
доцент кафедры фонетики и мето-
дики преподавания иностранных
языков

Санкт-Петербургский государст-
венный университет
г. Санкт-Петербург, Россия

Vera Evdokimova

PhD in Philology, Associate
professor of the Department of
Phonetics and Methods of Teaching
Foreign Languages

Saint Petersburg State University
Saint Petersburg, Russia
v.evdokimova@spbu.ru

Максимова Мария Романовна

преподаватель
Санкт-Петербургский
государственный университет
г. Санкт-Петербург, Россия

Maria Maksimova

Teacher
Saint Petersburg State University
Saint Petersburg, Russia
maromax@yandex.ru

**ФОНЕТИЧЕСКИЕ И МИМИЧЕСКИЕ ХАРАКТЕРИСТИКИ РЕЧИ
ПРИ КОГНИТИВНОЙ НАГРУЗКЕ**

В данной работе рассмотрены речевые и мимические характеристики когнитивной нагрузки. Материалами для исследования послужили аудио- и видеозаписи речи информантов во время прохождения когнитивного теста и симулятора вождения с дополнительными заданиями. Также в качестве материала использовались видеофрагменты шоу «50 вопросов» (интервью в машине, вопросы задает водитель). Акустические признаки были извлечены с помощью Praat. Мимические параметры (двигательные единицы по системе кодирования лицевых движений (СКЛиД)) были найдены с помощью OpenFace 2.2.0. Было выявлено, что при когнитивной нагрузке возрастает длительность отдельных звуков, повышается частотность восходящих интонационных контуров, сужается диапазон ЧОТ, увеличивается длительность ларингализованных участков. Результаты показали, что когнитивная нагрузка чаще всего проявляется в мимических движениях в области глаз и губ. Наиболее частотными и интенсивными лицевыми движениями при когнитивной нагрузке были ямочка (сокращение щечной мышцы), подниматель верхней губы и опускающий брови. Были определены особенности синхронизации речевых и мимических характеристик. В большинстве случаев фонетические и мимические проявления когнитивной нагрузки совпадают по времени начала.

Ключевые слова: фонетика; когнитивная нагрузка; акустический анализ речи; система кодирования лицевых движений (СКЛиД); обработка мультимодальных данных.

PHONETIC AND VISUAL CHARACTERISTICS OF COGNITIVE LOAD

This paper analyses phonetic features of speech under cognitive load, without cognitive load, and in final lengthening. To study speech under cognitive load, the participants were presented the Stroop Colour and Word test. As a dual task, the subjects answered questions related to their field of studies. To analyse speech without significant cognitive load, read speech samples were recorded. The findings showed that under cognitive load the duration of vowels, consonants and laryngealised segments increased while fundamental frequency range became narrower. Two experiments were conducted in order to study speech and facial movements

under cognitive load. During the first experiment, the participants were asked to play a driving simulator game and simultaneously answer general knowledge questions. Audio and videosamples were recorded. The videos were processed using OpenFace 2.2.0. The information about action units (facial muscles movements) was obtained and analysed. The results showed that cognitive load is most often manifested in facial movements in the eye and lip areas. The most frequent and intensive facial movements under cognitive load were dimpler, upper lip raiser and eyebrow lowerer. The synchronization features of speech and facial characteristics were determined. In most cases, phonetic and visual manifestations of cognitive load started at the same time.

Key words: phonetics; cognitive load; acoustic analysis of speech; facial action coding system (FACS); multimodal data processing.

Когнитивная нагрузка отражает объем ресурсов, затрачиваемых рабочей памятью человека при выполнении какой-либо задачи. Рабочая память обеспечивает хранение и обработку информации. Так как объем и возможности рабочей памяти человека ограничены, для эффективного выполнения задачи необходимо контролировать нагрузку на рабочую память [1].

В современности распознавание когнитивной нагрузки позволяет сделать голосовые помощники более удобными для использования. Это особенно важно для повышения безопасности систем, управляемых операторами (например, систем управления воздушным движением). При обнаружении признаков умственной нагрузки в речи пользователя система сможет адаптироваться под его психическое и физиологическое состояние [2].

При когнитивной нагрузке увеличивается напряжение голосовых складок. В результате, повышается подсвязочное давление. Это приводит к повышению частоты основного тона (ЧОТ) [3]. Кроме того, о наличии когнитивной нагрузки может свидетельствовать повышение интенсивности голоса [3; 4], сужение диапазона ЧОТ [2; 3; 4], ларингализация [2; 3; 5], изменение темпа речи и скорости артикуляции [там же].

Исследования показали, что значения формант, темп речи и скорость артикуляции могут как повышаться, так и понижаться при когнитивной нагрузке. Изменения темпа речи и скорости артикуляции могут быть связаны с заданиями, которые выполняются информантами, или с условиями эксперимента. Например, в исследовании К. Мюллера [6] было выявлено, что ограничение по времени приводит к повышению скорости артикуляции. Тем не менее при наличии дополнительной задачи скорость артикуляции, как правило, понижается. Эксперимент Т. Ф. Япа [5] показал, что при высокой когнитивной нагрузке темп речи возрастает, однако низкая и средняя нагрузка не вызывает изменений темпа речи. Как показал перцептивный эксперимент Ф. Н. Ли [3], средний уровень нагрузки в речи определяется на слух по медленному и прерывистому темпу речи, а высокая когнитивная нагрузка – по быстрому, но равномерному темпу.

Исследования мимики важны для распознавания когнитивной нагрузки, потому что психическое и физическое состояние человека можно определить, в том числе, по выражению лица, положению головы и движению глаз [7].

В работе А. Хавыло и его соавторов [8] утверждается, что психические состояния человека можно распознавать по произвольным движениям, таким, как «микродвижения неконтролируемых мимических мышц». В ходе эксперимента каждому из информантов задавалось 100 вопросов, из которых 50 вопросов были «когнитивно сложными», а остальные 50 вопросов были «когнитивно простыми». Была проведена автоматическая идентификация мимических микродвижений. Далее был выполнен статистический анализ полученных данных. Для построения модели распознавания когнитивной нагрузки использовались следующие характеристики: время обдумывания ответа на вопрос, время ответа, дисперсия мимических движений до ответа и дисперсия мимических движений во время ответа. Средняя точность определения уровня субъективной когнитивной нагрузки с помощью созданной модели составила 0,957. Следовательно, лицевые характеристики могут считаться надежным показателем для распознавания когнитивной нагрузки.

Также отмечается, что перед ответом на каждый вопрос у испытуемых наблюдалось изменение амплитуды мимических движений, которое не всегда было связано с началом артикуляции. Авторы заключают, что этот период мимической активности перед ответом является показателем когнитивной нагрузки или эмоциональной значимости стимула для информанта [9].

Одной из систем, позволяющих описывать и классифицировать выражения лица для их дальнейшей цифровой обработки, является система кодирования лицевых движений (СКЛиД) (англ. Facial Action Coding System (FACS)) [10]. Эта система характеризует выражение лица с помощью двигательных единиц (Action Units (AU)). Двигательные единицы обозначают движения мышц, формирующие определенное выражение лица. Система FACS широко используется для распознавания психического и физического состояния человека по выражению лица, в частности для распознавания когнитивной нагрузки и эмоций [7].

В настоящее время визуальные данные могут использоваться для повышения эффективности распознавания речи. Например, Д. Иванько и его коллеги создали мультимодальный корпус речи RUSAVIC (Russian Audio-Visual Speech in Cars) [11]. Этот корпус делает возможным распознавание речи в шуме с помощью автоматического чтения по губам. Данные для корпуса записывались при различных уровнях шума. Аудиоданные включают в себя наиболее частотные голосовые запросы в навигаторе. Также информанты произносили названия букв русского алфавита и наименования чисел.

Распознавание речи водителей необходимо, поскольку ручное управление навигатором может мешать следить за ситуацией на дороге. Однако определение физиологического и психического состояния водителей (включая когнитивную нагрузку) также является важной задачей. Вероятность человеческой ошибки, которая является частой причиной автомобильных

аварий, возрастает по мере снижения когнитивных способностей. В числе факторов, влияющих на когнитивные ресурсы, можно назвать стресс, усталость и другие психологические и физические состояния.

Целью первого эксперимента было изучение характеристик речи при когнитивной нагрузке. Исследование выполнялось на материале студийных аудиозаписей речи семи студенток кафедры фонетики и методики преподавания иностранных языков СПбГУ. Задание, вызывающее когнитивную нагрузку, было основано на тесте Струпа [12]. Тест Струпа – это нейропсихологический тест, позволяющий оценить способность человека подавлять реакцию на один из стимулов, если это затрудняет обработку второго стимула (оба стимула подаются испытуемому одновременно). Информантам было предъявлено 20 названий цветов, где цвет шрифта отличался от значения слова. Испытуемые должны были назвать цвет шрифта. В данном эксперименте подавляются вербальные функции, ответственные за чтение слова и понимание его значения. Это необходимо, так как вербальные функции затрудняют выполнение сенсорно-перцептивных функций, обеспечивающих распознавание цвета шрифта. Дополнительным заданием были вопросы, связанные с фонетикой английского языка. Речь в случае отсутствия когнитивной нагрузки была реализована информантами при чтении текста.

Был проведен акустический анализ полученных записей с помощью программы Praat. Значения ЧОТ и длительностей были вычислены с помощью скриптов из коллекции SpeCT (The Speech Corpus Toolkit for Praat). Было проведено сравнение длительностей безударных гласных, щелевых согласных, сонантов и ларингализованных участков при наличии и отсутствии когнитивной нагрузки.

Результаты обработки аудиоматериала представлены в Таблицах 1–4.

Таблица 1

Длительность ударных гласных при наличии
и отсутствии когнитивной нагрузки

	Under cognitive load (ms)	Without cognitive load (ms)
/i/	87	67
/a/	119	82
/o/	96	51

Таблица 2

Длительность безударных гласных при наличии
и отсутствии когнитивной нагрузки

	Under cognitive load (ms)	Without cognitive load (ms)
/i/	132	78
/a/	178	68
/i/	331	65

Таблица 3

Длительность сонантов при наличии и отсутствии когнитивной нагрузки

	Under cognitive load (ms)	Without cognitive load (ms)
/m/	119	99
/n/	207	52

Таблица 4

Длительность щелевых согласных при наличии
и отсутствии когнитивной нагрузки

	Under cognitive load (ms)	Without cognitive load (ms)
/s/	201	89
/s'/'	324	146
/ž/	175	71

Результаты показали, что длительность безударных гласных в случае затруднения при выполнении задачи превышает длительность безударных гласных при отсутствии размышлений, в среднем, на 140 мс. Однако эти различия были статистически значимыми только для гласной /i/ в заударных частях фонетических слов. Длительность ударных гласных также увеличивается при когнитивной нагрузке (в среднем, на 36 мс). Как и в случае безударных гласных, результаты не всегда были статистически значимыми. Существенное различие было обнаружено только в отношении гласной /a/.

Длительность щелевых согласных при когнитивной нагрузке также была выше, чем при отсутствии затруднений (в среднем, на 120 мс). Аналогично, средняя длительность сонантов при когнитивной нагрузке, в среднем, на 85 мс превышала длительность сонантов при отсутствии нагрузки. Тем не менее эти результаты не были статистически значимыми.

Длительность ларингализованных участков при когнитивной нагрузке была значимо больше, чем в случае предпаузного удлинения.

В случае затруднений при выполнении задания, как и при чтении неконечных синтагм в фонетически представительном тексте дикторы в большинстве случаев реализовали восходящее движение ЧОТ. Также частотным был восходяще-нисходящий интонационный контур.

Диапазон значений ЧОТ в восходящих контурах при когнитивной нагрузке сужался по сравнению с диапазоном восходящего мелодического движения при чтении неконечных синтагм в фонетически представительном тексте (в среднем, на 4 полутона). Это означает, что когнитивная нагрузка может проявляться в меньшем разнообразии мелодики по сравнению с речью при отсутствии когнитивной нагрузки.

Увеличение длительности гласных и согласных при когнитивной нагрузке позволяет сделать вывод, что скорость артикуляции, в целом, понижается при данном условии.

Целью второго эксперимента было изучение как акустических, так и визуальных показателей когнитивной нагрузки, а также их взаимосвязи. В этом эксперименте приняли участие 16 испытуемых (6 мужчин, 10 женщин). Возраст участников – от 19 до 52 лет, средний возраст – 33 года.

Испытуемые проходили симулятор вождения по городу “City Car Driving” и одновременно отвечали на вопросы на общие знания (кругозор). В ответе на некоторые вопросы требовалось перечислить какие-либо объекты (например, «Какие города, названия которых начинаются на букву “Р”, вы знаете?»). Другие вопросы предполагали более развернутый ответ («Опишите круговорот воды в природе»).

Было записано 16 видеофрагментов (по одной записи для каждого участника) с помощью смартфона с разрешением экрана FullHD 1920x1080. Продолжительность записей варьировалась от 7 до 13 минут. Общая продолжительность материала составила 2 часа 46 минут.

Полученные видеозаписи были обработаны с помощью OpenFace [13] (версия 2.2.0). Из записей были извлечены значения частоты и интенсивности двигательных единиц по СКЛиД. Мимические жесты в видеороликах также были аннотированы вручную с помощью ELAN, поскольку в некоторых случаях область лица не была корректно определена с помощью OpenFace. Автоматическая транскрипция аудиоматериала была выполнена с помощью WEBMAUS Basic. Ручная коррекция полученной транскрипции была проведена в программе Praat. Акустические параметры были получены с помощью скриптов SpeCT (The Speech Corpus Toolkit для Praat).

Результаты ручной аннотации, показали, что когнитивные трудности при ответе на вопросы чаще всего проявлялись в отведении глаз вправо вверх (AU62+AU63) или вниз вправо (AU61+AU63). В большинстве случаев отведение глаз вправо сопровождалось поворотом головы в том же направлении.

Частотные значения единиц действия, найденные с помощью OpenFace, приведены в Таблице 5.

Таблица 5
Двигательные единицы с наибольшей частотностью (симулятор вождения)

Номер ДЕ	Название ДЕ	Средняя частотность, %
AU14	Ямочка	66
AU05	Подниматель верхнего века	41
AU04	Опускатель брови	39
AU10	Подниматель верхней губы	37
AU45	Моргание	36
AU25	Разведение губ	31
AU02	Подниматель внешней части брови	24

Как показано в Таблице 5, когнитивная нагрузка, как правило, сопровождается движениями в области глаз и губ. Самым частотным жестом было сокращение щечной мышцы (ямочка). Что касается AU25 (разведение губ), это движение может быть связано с артикуляцией, и поэтому возможно, что оно не является надежным показателем когнитивной нагрузки.

Были отмечены различия между мужчинами и женщинами по движениям AU25 (разведение губ) и AU45 (моргание). У женщин эти двигательные единицы имели более высокую относительную частотность.

Двигательная единица AU45 встречалась среди мимических жестов девяти из десяти испытуемых женского пола. Для трех из них это действие было наиболее частотным, для двух женщин оно было на втором месте по частотности, для четырех испытуемых женского пола оно было на третьем месте.

Движение AU25 наблюдалось в мимике пяти информанток. Для трех из них это движение было вторым по частотности. Для других двух участниц оно занимало третье место по частотности.

У испытуемых мужского пола относительная частотность AU45 и AU25 варьировалась от 5-го до 12-го места.

Таблица 6

Двигательные единицы с наибольшей интенсивностью
(симулятор вождения)

Номер ДЕ	Название ДЕ	Средняя интенсивность, %
AU07	Натягиватель века	0,912
AU14	Ямочка	0,846
AU10	Поднятие верхней губы	0,659
AU04	Опускатель брови	0,621
AU25	Губы разведены	0,566
AU17	Подниматель подбородка	0,512
AU26	Челюсть опущена	0,464

Как показано в Таблицах 5 и 6, некоторые движения лица (а именно, AU14, AU10, AU04 и AU25) демонстрируют относительно высокие значения как по частоте, так и по интенсивности. Сокращение щечной мышцы (AU14) было на втором месте по частоте и на втором месте по интенсивности. Поднятие верхней губы (AU10) было четвертой по частоте двигательной единицей и также имело относительно высокое значение интенсивности (третье место). Аналогичная ситуация наблюдается для опускателя брови (AU04). Напротив, двигательная единица AU05 (подниматель верхнего века) была достаточно частотной, но показала относительно низкую интенсив-

ность. Кроме того, наиболее интенсивное движение, AU07 (натягиватель века), имело низкую частотность. Таким образом, интенсивность двигательных единиц не всегда коррелирует с частотностью.

Из обеих таблиц следует, что когнитивную нагрузку можно определить на основе мимических движений в области глаз и губ. Ямочка, поднятие верхней губы и опускание брови, по-видимому, являются наиболее значимыми характеристиками когнитивной нагрузки, так как имеют высокую частотность и интенсивность.

У всех испытуемых наблюдались случаи, когда показатели когнитивной нагрузки в речи и мимике проявлялись одновременно. Тем не менее частотность таких синхронных проявлений различалась. В частности, паузы, заполненные паузами хезитации, часто сопровождалось отведением глаз. Кроме того, многие информанты реализовывали восходящее движение ЧОТ, когда смотрели вверх. В этом случае значение ЧОТ и направление взгляда изменялись в одном и том же направлении (вверх).

В целом, наблюдались следующие временные сценарии для акустических и визуальных характеристик когнитивной нагрузки:

1. Акустические и визуальные маркеры (одновременно)
2. Акустические маркеры предшествуют визуальным
3. Визуальные маркеры предшествуют акустическим
4. Только акустические маркеры
5. Только визуальные маркеры

Первый сценарий (одновременное появление акустических и визуальных характеристик) был наиболее частотным для одиннадцати участников из шестнадцати. В записях этих дикторов когнитивная нагрузка была выражена одновременно в речи и в мимике в 33 %–50 % от общего числа сегментов с когнитивной нагрузкой. Случай, в котором акустические маркеры наблюдаются раньше визуальных, был вторым по частотности.

Средняя продолжительность сегментов, в которых одновременно присутствуют акустические и визуальные характеристики когнитивной нагрузки, составила 1,8 секунды. Это на 0,43 секунды больше, чем средняя продолжительность сегментов, в которых наблюдались только визуальные параметры. Это наблюдение может быть объяснено увеличением длительности звуков и высокой частотой заполненных пауз хезитации в сегментах с одновременным возникновением акустических и визуальных параметров когнитивной нагрузки.

Целью третьего эксперимента было изучение акустических и мимических характеристик речи при вождении и одновременном выполнении речевой задачи. Материалом послужили три выпуска шоу «50 вопросов» (интервью в машине, вопросы задает водитель; некоторые фрагменты интервью проходят в студии) [14].

Все фрагменты, снятые в автомобиле, были отделены от сцен, снятых в студии. Из всех фрагментов, происходящих в машине, были вырезаны сцены, в которых говорит водитель. Общая продолжительность материалов составила 2 часа 42 минуты (разрешение экрана FullHD 1920x1080, формат mp4).

Полученные видеозаписи были обработаны с использованием OpenFace 2.2.0. Как и в эксперименте с симулятором вождения, из видеозаписей были извлечены двигательные единицы. Выражение лица водителя при разговоре сравнивалось с выражением его лица во всех сценах в автомобиле в целом. Сравнивались его речевые характеристики при проведении интервью в машине и в студии.

Из видео были извлечены звуковые дорожки. Автоматическая транскрипция звуковых файлов была выполнена с помощью WEBMAUS Basic. Полученные орфографические и фонетические изменения были скорректированы вручную с помощью Praat. Акустические параметры были получены с помощью скриптов SpeCT (The Speech Corpus Toolkit for Praat).

Найденные длительности звуков в речи водителя при одновременном вождении (то есть, при повышенной когнитивной нагрузке) и в студии (без дополнительной когнитивной нагрузки) представлены в Таблицах 7–9.

Таблица 7

Длительность ударных гласных
при одновременном вождении и в студии

	Вождение и речь (ms)	Только речь (ms)
/a/	214	85
/o/	302	82
/i/	236	54

Таблица 8

Длительность безударных гласных
при одновременном вождении и в студии

	Вождение и речь (ms)	Только речь (ms)
/a/	164	43
/i/	177	53
/i/	233	54

Таблица 9

Длительность сонантов и щелевых согласных
при одновременном вождении и в студии

	Вождение и речь (ms)	Только речь (ms)
/m/	187	56
/n/	174	42
/z/	213	74

В Таблицах 7–9 показано, что длительность гласных и согласных звуков увеличивалась при одновременном вождении автомобиля и разговоре. Средние различия между длительностями звуков в этих двух условиях были в целом выше, чем показали результаты эксперимента с использованием теста Струпа. При выполнении дополнительной задачи длительность произнесения ударных и безударных гласных была выше на 177 и 141 мс соответственно по сравнению с речью интервьюера в студии. Разница между значениями длительности сонантов в этих двух условиях составляла в среднем 132 мс. Что касается щелевых согласных, то только один звук этого типа (/z/) удлинялся при наличии высокой когнитивной нагрузки.

Наиболее частотные двигательные единицы в мимике во всех сценах вождения перечислены в Таблице 10. Наиболее частотные двигательные единицы при одновременном выполнении речевой задачи указаны в Таблице 11.

Таблица 10

Частотность двигательных единиц
во всех сценах вождения

Номер ДЕ	Название ДЕ	Средняя частотность, %
AU14	Ямочка	72
AU05	Подниматель верхнего века	54
AU02	Подниматель внешней части брови	52
AU01	Подниматель внутренней части брови	45
AU15	Опускатель уголка губы	42
AU10	Подниматель верхней губы	41

Таблица 11

Частотность двигательных единиц
при вождении и выполнении речевой задачи

Номер ДЕ	Название ДЕ	Средняя частотность, %
AU14	Ямочка	64
AU02	Подниматель внешней части брови	51
AU10	Подниматель внешней части брови	49
AU05	Подниматель верхнего века	46
AU01	Подниматель внутренней части брови	44
AU20	Растягиватель губ	38

Как показано в Таблицах 10 и 11, наиболее частой двигательной единицей в обеих ситуациях была AU14 (72 % при управлении автомобилем в целом и 64 % при одновременном управлении автомобилем и разговоре). Другими двигательными единицами с относительно высокой частотой в обоих этих условиях были AU05, AU01 и AU02, которые представляют собой движения в области глаз. По-видимому, нет существенных различий в относительной частоте различных двигательных единиц в Таблицах 10 и 11. Однако выполнение двойной задачи в целом привело к несколько более низким значениям частоты. Это может означать, что более высокая когнитивная нагрузка приводит к снижению мимической активности.

Димплер (AU14) также показал самую высокую частоту в эксперименте с симулятором. Частота использования этого элемента управления в первом эксперименте была аналогична частоте его использования при разговоре в реальных условиях вождения (66 % и 64 % соответственно). Тем не менее между результатами этих двух экспериментов наблюдаются некоторые различия. Что касается фрагментов шоу, в которых водитель проводил интервью за рулем, то показатели AU01 и AU02 (внутреннее и внешнее поднятие бровей) имели относительно высокие значения частоты – 44 % и 51 % соответственно. Эти значения примерно в два раза превышают частоты этих действий в эксперименте на симуляторе, которые составляли 21 % и 24 % соответственно. Поскольку вождение автомобиля в реальных условиях, вероятно, вызывает более высокую когнитивную нагрузку, чем игра в симуляторе, это различие может указывать на то, что поднятие бровей может служить показателем более значительной когнитивной нагрузки, несмотря на общее снижение мимической активности.

Значения интенсивности двигательных единиц во всех сценах вождения и конкретно при выполнении речевой задачи представлены в Таблицах 12 и 13.

Таблица 12

Двигательные единицы с наибольшей интенсивностью
(все сцены вождения)

Номер ДЕ	Название ДЕ	Средняя частотность, %
AU14	Ямочка	1,401086
AU10	Подниматель верхней губы	1,027073
AU25	Губы разведены	0,840993
AU17	Подниматель подбородка	0,799847
AU26	Челюсть опущена	0,706391
AU04	Опускание брови	0,680125

Таблица 13

Двигательные единицы с наибольшей интенсивностью
(речь при вождении)

Номер ДЕ	Название ДЕ	Средняя частотность, %
AU14	Ямочка	1,167815
AU10	Подниматель верхней губы	1,014202
AU17	Подниматель подбородка	0,611325
AU25	Губы разведены	0,541910
AU04	Опускание брови	0,529352
AU26	Челюсть опущена	0,488872

Как показано в Таблицах 12 и 13, набор наиболее интенсивных двигательных единиц был одинаковым для управления автомобилем в целом и для разговора во время вождения. В обоих случаях эти действия были расположены относительно друг друга в порядке возрастания примерно одинаково. Единственное различие заключалось в порядке следования AU04 и AU26. В Таблице 12 AU26 предшествует AU04, тогда как AU04 предшествует AU26 в Таблице 13. При выполнении дополнительной речевой задачи интенсивность лицевых движений, как правило, была ниже по сравнению со всеми сценами вождения в целом. Это может свидетельствовать о том, что мимика становится менее интенсивной и ярко выраженной при более высоких уровнях когнитивной нагрузки.

В данном исследовании были изучены фонетические и визуальные характеристики когнитивной нагрузки и их взаимосвязь. Результаты показали, что при когнитивной нагрузке длительность как гласных, так и

согласных увеличивается, в то время как диапазон F0 сужается. Ларингализованные сегменты также имеют более высокую длительность при когнитивной нагрузке. При одновременном разговоре и вождении автомобиля (то есть при более высокой когнитивной нагрузке) длительность звуков увеличивалась больше, чем в эксперименте с тестом Струпа. Это может свидетельствовать о прямой зависимости между длительностью звуков и уровнем когнитивной нагрузки.

Что касается визуальных параметров, то было обнаружено, что когнитивную нагрузку можно распознать по мышечным движениям в области глаз и губ. Ямочка (сокращение щечной мышцы), поднятие верхней губы и опускание бровей имели относительно высокие значения частоты и интенсивности в обоих экспериментах. В большинстве случаев ямочка была самым частотным и интенсивным показателем. Это может означать, что эта двигательная единица является одним из наиболее достоверных показателей когнитивной нагрузки.

Однако результаты двух экспериментов в некоторых отношениях отличались друг от друга. Поднятие внутренней и внешней части бровей наблюдалось гораздо чаще в реальных условиях вождения при выполнении дополнительной речевой задачи по сравнению с экспериментом на симуляторе вождения. Следовательно, эти мимические жесты могут быть характерны для более высокого уровня когнитивной нагрузки. Анализ поведения одного и того же диктора (водителя) при разных уровнях когнитивной нагрузки показал, что интенсивность и частота двигательных единиц, как правило, понижаются при увеличении нагрузки. Это может означать, что повышенная когнитивная нагрузка может привести к снижению мимической активности.

Частота AU25 (часть губ) и AU45 (моргание) различалась у информантов разного пола.

Была изучена синхронизация фонетических и визуальных характеристик когнитивной нагрузки. Было выявлено, что визуальные и речевые маркеры чаще всего встречаются одновременно. Кроме того, средняя продолжительность одновременного появления различных показателей когнитивной нагрузки была больше, чем средняя продолжительность фрагментов, содержащих либо только фонетические, либо только визуальные параметры.

ЛИТЕРАТУРА

1. Sweller J. Cognitive load during problem solving: Effects on learning // Cognitive Science. 1988. № 12 (2). P. 257–285.
2. Berthold A., Jameson A. Interpreting Symptoms of Cognitive Load in Speech Input // Springer, Vienna. 2010. P. 235–244.
3. Le P. N. The Use of Spectral Information in the Development of Novel Techniques for Speech-Based Cognitive Load Classification: PhD thesis. Sydney, 2012.

4. Huttunen K., Keränen H., Väyrynen E., Pääkkönen R., Leino T. Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights // *Applied ergonomics*. 2011. № 42 (2). P. 348–357.
5. Yap T. F. Speech Production Under Cognitive Load: Effects and Classification: PhD thesis. Sydney, 2012.
6. Müller C., Großmann-Hutter B., Jameson A., Rummer R., Wittig F. Recognizing time pressure and cognitive load on the basis of speech: An experimental study // *International Conference on User Modeling*. Berlin, Heidelberg. 2001. P. 24–33.
7. Шишов Б. А. Оценка когнитивной загруженности оператора на основе видеоинформации с помощью рекуррентных нейронных сетей // *Проблемы управления*. 2018. № 2. С. 58–65.
8. Khavylo, A., Engalychev, V., Leonova, E., Danshin, V., Mirzeabasov, O.: Manifestation of Task's Cognitive Complexity in Mimic Micromovements: Prognostic Model. In: Arai, K. (eds) *Proceedings of the Future Technologies Conference (FTC) 2021*. Volume 2. *Lecture Notes in Networks and Systems*. Vol. 359. Springer, Cham.
9. Leonova E., Engalychev V., Khavylo A., Mirzeabasov O., Danshin V. Mimic indicators of task complexity: individual approach // *14th Conference of the European Human Behavior and Evolution Association*. 2019. P. 199–200.
10. Ekman P., Friesen W. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto : Consulting Psychologists Press, 1978.
11. Ivanko D., Ryumin D., Axyonov A., Kashevnik A., Karpov A.: Multi-Speaker Audio-Visual Corpus RUSAVIC: Russian Audio-Visual Speech in Cars. In: *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pp. 1555–1559. European Language Resources Association (ELRA), Marseille (2022).
12. Stroop J. R. Studies of interference in serial verbal reactions // *Journal of Experimental Psychology*. 1953. № 18 (6). P. 643–662.
13. Baltrusaitis T., Zadeh A., Lim Y. C., Morency L.-P. OpenFace 2.0: Facial Behavior Analysis Toolkit // *13th IEEE International Conference on Automatic Face & Gesture Recognition*. 2018. P. 59–66.
14. Макарена (50 вопросов) [Electronic resource] // URL: <https://vk.com/makarena50q> (date of access: 01.07.2025).

УДК 81'33

Зиновьева Анастасия Юрьевна
кандидат филологических наук,
доцент
Южно-Уральский государственный
университет
г. Челябинск, Россия

Anastasiia Zinoveva
PhD in Philology, Associate Professor
South Ural State University
Chelyabinsk, Russia
zinovevaai@susu.ru

ЭМОДЗИ В АНАЛИЗЕ ТОНАЛЬНОСТИ (на материале отзывов о медцентрах на русском языке)

Цель исследования – выявить особенности функционирования эмодзи в интернет-отзывах о медицинских центрах Челябинска и оценить их применимость в многоаспектном анализе тональности, в частности определить влияние эмодзи на вычисление показателя тональности. В качестве материала исследования использован корпус из 3247 отзывов пользователей электронного справочника 2GIS о 22 медицинских центрах города (по состоянию на март 2023 г.) общим объемом 200 265 словоупотреблений. В работе использованы количественные и качественные методы анализа: корпусный и контекстный анализ, лексико-онтологическое моделирование, сравнительный анализ. Инструментарий исследования включает базу данных эмодзи Emoji Sentiment Ranking, а также инструменты НОЦ «Лингво-инновационные технологии» ЮУрГУ для многоаспектного анализа тональности: онтологию предметной области, русскоязычный онтолексикон, платформу концептуального аннотирования, «золотой» корпус. Результаты исследования продемонстрировали, что в рассмотренном корпусе эмодзи используются для выражения различных градаций оценки и усиления тональности, для передачи оценок и (значительно реже) аспектов, могут дублировать тональность высказывания или выступать самостоятельным индикатором тональности.

Ключевые слова: эмодзи; многоаспектный анализ тональности; тональная классификация эмодзи; интернет-отзывы; русский язык; медицинские услуги.

EMOJIS IN SENTIMENT ANALYSIS: A CASE STUDY IN RUSSIAN MEDICAL CENTER REVIEWS

The study aims to identify the functional characteristics of emojis in online reviews of medical centers in Chelyabinsk, and to assess their potential use in multi-aspect sentiment analysis—specifically, how emojis influence sentiment score calculations. The research material is a corpus of 3,247 user reviews extracted from the 2GIS electronic database, covering 22 city medical centers as of March 2023, with a total word count of 200,265. The methods employed in this study include quantitative and qualitative approaches, such as corpus and contextual analysis, lexical-ontological modeling, and comparative analysis. The research tools encompass the Emoji Sentiment Ranking database and multi-aspect sentiment analysis tools developed by the Innovation in Linguistics Research and Education Center at South Ural State University: a domain ontology, a Russian-language ontollexicon, a conceptual annotation platform, and a gold-standard corpus. The findings indicate that, within the corpus under analysis, emojis are used to express and intensify various sentiments, convey opinions or (much less frequently) evaluated aspects; they may reproduce the sentiment expressed by the text they accompany or act as a standalone sentiment indicator.

Key words: emoji; multi-aspect sentiment analysis; emoji sentiment classification; online reviews; Russian; medical services.

В настоящее время одним из наиболее перспективных направлений в области автоматической обработки текста является анализ тональности, поскольку он позволяет, среди прочего, извлекать из интернет-отзывов ценные сведения о мнениях потребителей тех или иных услуг. В этом контексте особое значение приобретают **эмодзи** (от яп. э ‘картинка’ + *модзи* ‘знак’) – «пиктограммы, используемые в интернет-коммуникации и изображающие лица, животных, распространенные предметы, знаки и природные явления» [1, с. 233], которые представляют собой значимые невербальные элементы, способные дополнять или модифицировать смысл исходного сообщения. Учет эмодзи в анализе тональности позволяет получать более точные результаты [2–4], особенно при обработке коротких сообщений, где эмодзи могут быть единственным индикатором настроения автора [5]. При этом в отличие от эмотиконов (от англ. *emotion* ‘эмоция’ + *icon* ‘значок’) эмодзи могут быть использованы не только для передачи эмоций, но и для обозначения объектов, действий и идей, что потенциально позволяет извлекать из сообщений, помимо оценок и мнений, и сами оцениваемые аспекты.

Цель настоящей работы – выявить особенности функционирования эмодзи в корпусе интернет-отзывов об услугах частных медицинских центров города Челябинска на русском языке и оценить степень их применимости в многоаспектном анализе тональности, в частности определить их влияние на вычисление показателя тональности фрагментов данного корпуса.

Работа выполнена в рамках проекта, посвященного многоаспектному анализу тональности отзывов о медицинских центрах Челябинска, который реализуется в НОЦ «Лингво-инновационные технологии» Южно-Уральского государственного университета. В проекте используются специально разработанные предметно-ориентированные лингвистические ресурсы, часть которых задействована в настоящем исследовании, а именно:

1) **исходный корпус**: 3247 интернет-отзывов пользователей электронного справочника 2GIS [6] о 22 медицинских центрах города (по состоянию на март 2023 г.) общим объемом 200 265 словоупотреблений;

2) **онтология предметной области**, ориентированная на многоаспектный анализ тональности, которая содержит четыре ветви концептов: ASPECT (аспект оценки: например, Q – качество оказания услуг и результат лечения, INT – интерьер медицинского центра, его дизайн и оснащение), SENTIMENT (тональность: POS – положительная, POSX – крайне положительная, NEG – отрицательная, NEGX – крайне отрицательная), MODIFIER (модификатор тональности: ANT – ослабление, MAG – усиление, REV – изменение на противоположную), OTHER;

3) **исходный онтолексикон** – список релевантных лексических единиц, извлеченных из корпуса, соотнесенных с концептами онтологии;

4) **платформа концептуального аннотирования** [7]: теггер на основе лексико-онтологической базы знаний предметной области, который обеспечивает автоматическую разметку текста аспектными и тональными метками с возможностью последующего постредактирования разметки пользователем;

5) **размеченный корпус** отзывов о медицинских центрах с эталонной аспектно-тональной разметкой («золотой» корпус) без учета эмодзи (подробнее о разработке корпуса см. [8]).

Исследование было проведено в четыре этапа.

На **первом этапе** из исходного корпуса были извлечены все эмодзи с левым и правым контекстом в размере предложения. Каждому эмодзи были поставлены в соответствие код и название по стандарту Unicode [9]; например, эмодзи 🍑 соответствует код U+1F44D и название «thumbs up». Эмодзи, обозначающие людей или части тела, как правило, представлены в нескольких вариантах, которые различаются цветом кожи и (или) половой принадлежностью изображенных людей; в этом случае они имеют дополнительные коды. При подсчете частоты встречаемости эмодзи с общим основным кодом были объединены, поскольку указанные различия представляются нерелевантным для поставленной задачи.

На **втором этапе** стояла задача классифицировать каждый эмодзи. В контексте анализа тональности эмодзи традиционно подразделяют на положительные, отрицательные и нейтральные, однако возможна более подробная классификация, если это продиктовано задачами исследования [10]. В нашем случае за основу классификации была взята описанная выше онтология. Кроме того, для определения тональности эмодзи в выделенных фрагментах отзывов применялись база данных Emoji Sentiment Ranking [11] и контекстный анализ эмодзи на основе данных «золотого» корпуса, что позволило отнести каждый эмодзи к одному или нескольким концептам онтологии, принадлежащим ветвям ASPECT, SENTIMENT, MODIFIER и OTHER.

На **третьем этапе** исследования была произведена лексикализация эмодзи в выделенных фрагментах исходного корпуса, т. е. каждый эмодзи был заменен на его название по Unicode; также соответственно был обновлен онтолексикон. Далее фрагменты исходного корпуса были повторно размечены на платформе концептуального аннотирования с обновленной лексико-онтологической базой знаний и подвергнуты последующему постредактированию для снятия неоднозначности меток, в результате чего был создан «золотой» корпус с учетом эмодзи.

На **четвертом этапе** были рассчитаны и сопоставлены друг с другом показатели тональности каждого фрагмента «золотого» корпуса без учета и с учетом эмодзи. Показатель тональности (S) рассчитывался по формуле:

$$S = \frac{1}{n} \sum_{i=1}^n c_i m_i + 3, (1)$$

где n – количество тонально окрашенных единиц в тексте;
 c_i – коэффициент тональности i -й единицы;
 m_i – коэффициент, учитывающий влияние модификатора.

Примечание: коэффициент c_i принимает следующие значения: POSX = 2, POS = 1, NEG = -1, NEGX = -2. Коэффициент m_i равен 2 при воздействии на тонально окрашенную единицу модификатора MAG (кроме случаев, когда такая единица имеет тональность POSX или NEGX, для которых $m_i = 1$); 0,5 при воздействии модификатора ANT и 1 в остальных случаях. Показатель тональности S принимает значения в диапазоне от 1 до 5, где 1 соответствует максимально отрицательной тональности, а 5 – максимально положительной. Если $n = 0$, показатель тональности не вычисляется.

По результатам первого этапа исследования в рассматриваемом корпусе обнаружено 286 фрагментов, содержащих от одного до десяти эмодзи, причем в 285 случаях эмодзи иллюстрируют левый контекст и лишь в одном – правый. Общее количество эмодзи в корпусе составило 402. Среди них выделено 78 уникальных эмодзи (при этом эмодзи, отличающиеся только цветом кожи или половой принадлежностью изображенных людей, но имеющие общий код по стандарту Unicode, учитывались как один символ), 28 из которых встречаются в корпусе только один раз. Десять наиболее частых эмодзи представлены в Таблице 1.

Таблица 1

Частотный рейтинг десяти наиболее употребимых эмодзи
в исследуемом корпусе отзывов о медицинских центрах

№ п/п	Эмодзи	Код по Unicode	Название по Unicode	Частота
1		U+1F44D	thumbs up	72
2		U+2764	red heart	29
3		U+1F64F	folded hands	25
4		U+1F44E	thumbs down	16
5		U+1F621	enraged face	15
6		U+1F60A	smiling face with smiling eyes	15
7		U+263A	smiling face	14
8		U+1F926	person facepalming	13
9		U+1F601	beaming face with smiling eyes	11
10		U+1F490	bouquet	11

На втором этапе эмодзи были распределены по девяти концептам онтологии, среди которых представлены как концепты, обозначающие тональность и ее модификаторы (POS, POSX, NEG, NEGX, MAG), так и концепты ветвей ASPECT и OTHER. При этом эмодзи, ослабляющие положительную или отрицательную тональность высказывания (ANT), а также изменяющие тональность на противоположную (REV), в корпусе не выявлены. Лишь три эмодзи были отнесены к концептам ветви ASPECT, что свидетельствует о предпочтении авторов отзывов выражать оцениваемые аспекты вербально.

Эмодзи, отнесенные к концепту OTHER, не обладают тональностью и не изменяют тональность высказывания вообще или в отдельных случаях употребления в корпусе, либо обозначаемая ими эмоция относится к самому автору, а не к аспектам работы медицинского центра. Например, в предложении *Мы обошли несметное количество врачей, которые разводили руки* 🙄 (примеры приводятся по исходному корпусу в авторской редакции) эмодзи 🙄 иллюстрирует выражение *разводили руки* и является нетональным, а в предложении *Хочется отметить медсестру из процедурной, (к огромному своему стыду не знаю имени 😞)* эмодзи 😞 передает сожаление автора по поводу незнания имени медсестры и не относится к оценке медицинского центра.

Некоторые эмодзи проявляют неоднозначность в корпусе, вследствие чего они были отнесены к нескольким концептам. Преимущественно это традиционно положительные эмодзи (👍, 😊, 🙌, 😄 и т. п.), которые в отдельных контекстах выражают иронию и сарказм. Например, в предложении *Краснуха, которая прошла после одной таблетки супрастина* 🙌👍 оба эмодзи несут иронический подтекст, отражающий сомнение автора по поводу поставленного диагноза.

Повторение эмодзи усиливает их тональность, поэтому группы из двух и более одинаковых эмодзи, имеющих тональность POS и NEG, отнесены соответственно к концептам POSX и NEGX. Частный случай – эмодзи ★, который, по данным Emoji Sentiment Ranking, преимущественно позитивен либо нейтрален, однако его также используют для выражения оценки по пятибалльной шкале. Соответственно, в таком контексте ★ соответствует концепту NEGX, ★★ – NEG, ★★★★★ – POS, ★★★★★★ – POSX (случай с тремя звездами потребует дополнительного контекстного анализа, поскольку оценка «три» по пятибалльной шкале может быть как нейтральной, так и отрицательной).

Итоговое распределение эмодзи по концептам приведено в Таблице 2.

Таблица 2

Распределение эмодзи по концептам онтологии

Концепт	Описание концепта	Эмодзи
POS	Положительная тональность; базовые положительные эмоции (радость, одобрение, благодарность)	
POSX	Крайне положительная тональность; более интенсивные положительные эмоции (восторг, восхищение, глубокое удовлетворение)	
NEG	Отрицательная тональность; базовые отрицательные эмоции (грусть, разочарование, злость, тревога); ирония и сарказм	
NEGX	Крайне отрицательная тональность; более интенсивные отрицательные эмоции (отчаяние, ярость, ужас); насмешка	
MAG	Усиление тональности высказывания	
OTHER	Эмодзи, не имеющие тональности и не изменяющие тональность высказывания либо не относящиеся к оценке медицинского центра	
INT	Интерьер медцентра, его дизайн, оснащение	
Q	Качество оказания услуг, результат лечения	

Результаты сравнительного анализа показателей тональности фрагментов в выборках «золотого» корпуса с учетом и без учета эмодзи приведены на Рисунке 1. На графике представлено распределение показателей тональности 242 из 286 фрагментов. Большая часть точек расположена вблизи пунктирной линии – это свидетельствует о том, что присутствие эмодзи не оказывает существенного влияния на тональность фрагмента. В тех случаях, когда влияние имеется, оно носит незначительный характер.

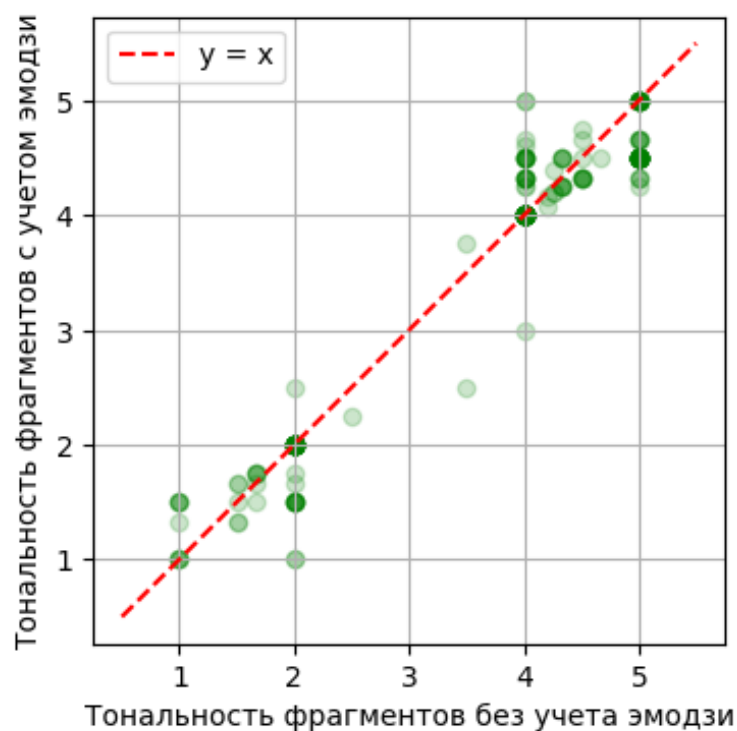


Рис. 1. Сравнение тональности фрагментов корпуса с учетом и без учета эмодзи

Показатель тональности 44 из 286 фрагментов изначально не рассчитывался из-за отсутствия тонально окрашенных лексических единиц, однако включение эмодзи позволило определить тональность еще 37 из них (оставшиеся семь фрагментов не имеют тональности, либо она нейтральна). Интересно, что в пяти фрагментах эмодзи выполняют функцию предиката и являются неотъемлемой частью высказывания, без которой невозможно определить его тональность; например: *Лор 🙌*; *Питание тоже 🙌*; *медсестры просто 🔥🔥🔥*. В остальных 32 фрагментах эмодзи выступают в роли дополнительного, факультативного элемента, уточняющего эмоциональный оттенок высказывания. Без эмодзи данные фрагменты остаются понятными и сохраняют базовую тональность, которую, однако, затруднительно определить с использованием тонального лексикона, поскольку оценка ситуации не выражена явно на уровне лексем. Примером является фрагмент *Мы нашли своего гастроэнтеролога 😊*, где ни одно из слов не обладает тональной окраской, однако совокупность слов формирует позитивную идею, подтверждаемую положительным эмодзи.

Влияние эмодзи на определение тональности текстовых фрагментов можно проиллюстрировать на примере самого распространенного в корпусе символа 🙌 *thumbs up* ‘палец вверх’. По данным *Emoji Sentiment Ranking*, эмодзи 🙌 в большей мере положителен: его показатель тональности

в диапазоне $[-1...1]$ равен 0,521. В рассматриваемом корпусе данный эмодзи, который встречается 72 раза в 57 фрагментах, был отнесен к концептам POS, NEG, OTHER или POSX (в группах от двух эмодзи), что отражено на Рисунке 2.

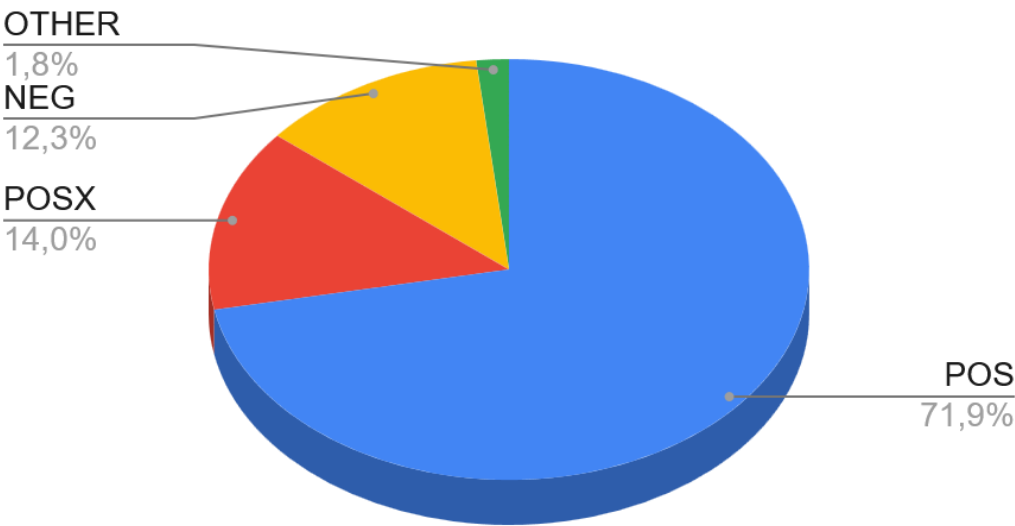


Рис. 2. Варианты тональности
эмодзи 👍 *thumbs up* ‘палец вверх’ в корпусе

Относительно влияния эмодзи 👍 на определение показателя тональности следует отметить, в 29 из 55 фрагментов показатель оставался неизменным; в 8 – изменялся менее чем на 0,5; в 14 – на 0,5; в 1 – на 0,67. В четырех фрагментах вычисление показателя тональности было возможно исключительно с учетом эмодзи. В одном случае включение эмодзи в расчет не повлияло на результат, поскольку эмодзи 👍 не обладал тональностью и иллюстрировал слово *ок*, использованное пациентом для описания своего состояния до обращения к врачу.

Итак, проведенный анализ корпуса отзывов о медицинских центрах позволил выявить несколько ключевых аспектов функционирования эмодзи и их влияния на определение тональности текста в рамках данной предметной области. Во-первых, проанализированные эмодзи служат средством выражения различных градаций оценки, включая положительную, крайне положительную, отрицательную и крайне отрицательную, а также способны усиливать тональность высказывания. Во-вторых, эмодзи могут применяться в многоаспектном анализе тональности не только для выявления оценок и мнений, но и для извлечения оцениваемых аспектов, однако в исследуемом корпусе наблюдается тенденция к вербальному выражению аспектов, эмодзи же преимущественно используются для передачи оценок. Наконец, преобладающая функция эмодзи заключается в дублировании тональности тексто-

вого фрагмента, однако они могут также выступать в качестве единственного индикатора тональности коротких фрагментов, что согласуется с результатами предыдущих исследований в данной области.

ЛИТЕРАТУРА

1. Словарь языка интернета.ru / Под. ред. М. А. Кронгауза. М.: АСТ-ПРЕСС КНИГА, 2016. 288 с.
2. Sentiment of Emojis [Electronic resource] / P. Kralj Novak, J. Smailović, B. Sluban, I. Mozetič // PLoS One. 2015. Vol. 10 (12). DOI: <https://doi.org/10.1371/journal.pone.0144296>.
3. EMFSA: Emoji-Based Multifeature Fusion Sentiment Analysis [Electronic resource] / H. Tang, W. Tang, D. Zhu, Sh. Wang, Y. Wang, L. Wang // PLoS One. 2024. Vol. 19 (9). DOI: <https://doi.org/10.1371/journal.pone.0310715>.
4. Alawneh H., Hasasneh A., Maree M. On the Utilization of Emoji Encoding and Data Preprocessing with a Combined CNN-LSTM Framework for Arabic Sentiment Analysis // Modelling. 2024. Vol. 5 (4). P. 1469–1489.
5. Emoji-Integrated Polyseme Probabilistic Analysis Model: Sentiment Analysis of Short Review Texts on Library Service Quality [Electronic resource] / X. Zheng, W. Chen, H. Zhou, Zhe Li, T. Zhang, Qi Yuan // IIETA. 2022. P. 313–322. DOI: <https://doi.org/10.18280/ts.390133>.
6. Карта Челябинска: улица, дома и организации города – 2GIS [Электронный ресурс]. URL: <https://2gis.ru/chelyabinsk> (дата обращения: 11.07.2025).
7. Шереметьева С. О., Бабина О. И. Платформа для концептуального аннотирования многоязычных текстов // Вестник ЮУрГУ. Серия «Лингвистика». 2020. Т. 17, № 4. С. 53–60.
8. Sheremetyeva S. O., Babina O. I. On Automated Creation of Gold-Standard Corpus for Multi-Aspect Sentiment Analysis // Internet and Modern Society. Human-Computer Communication. 2025 (in print).
9. Full Emoji List, v16.0 [Electronic resource]. URL: <https://unicode.org/emoji/charts/full-emoji-list.html> (date of access: 11.07.2025).
10. Hakami S. A. A., Hendley R., Smith Ph. Emoji Sentiment Roles for Sentiment Analysis: A Case Study in Arabic Texts // Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP). 2022. P. 346–355.
11. Emoji Sentiment Ranking v1.0 [Electronic resource]. URL: https://kt.ijs.si/data/Emoji_sentiment_ranking/index.html (date of access: 11.07.2025).

УДК 81.316.776.2

Керимова Самая Алифхан гызы
доктор философии по филологии,
доцент
Бакинский славянский университет
г. Баку, Азербайджан

Samayya Alifkhan gizi Karimova
Doctor of Philosophy in Philology,
Associate Professor
Baku Slavic University
Baku, Azerbaijan
samaya_kerimova@mail.ru

ИНТЕРНЕТ-МЕМЫ В ЦИФРОВОМ ПРОСТРАНСТВЕ: ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ

В статье рассматривается феномен интернет-мемов как важного средства цифровой культуры в современном обществе, показывается необходимость его изучения как способа межкультурной коммуникации и социализации. Анализируются возможности Интернет-мемов, которые обладают значительным потенциалом для повышения эффективности образовательного процесса, делая его актуальным и релевантным для современного поколения обучающихся. Основными выводами автора является определение роли Интернет-мема в современном цифровом пространстве, а также анализ его места в образовательной и культурной сфере. Обозначаются перспективы дальнейшего изучения и использования Интернет-мемов с учетом развития цифровых программ и новых платформ, алгоритмов распространения мемов и изменяющихся приоритетов пользователей.

Ключевые слова: цифровое пространство; интернет-мем; образовательная среда; аудиовизуальная культура; вербальные элементы; социокультурная компетенция; межкультурная коммуникация; коммуникативно-прагматический потенциал.

INTERNET MEMES IN DIGITAL SPACE: OPPORTUNITIES AND PROSPECTS

This article explores the phenomenon of internet memes as a significant element of digital culture in contemporary society, emphasizing the need to study them as a means of intercultural communication and socialization. It analyzes the potential of internet memes to enhance the effectiveness of the educational process by making it more relevant and engaging for today's generation of learners. The main conclusions of the author include defining the role of internet memes in the modern digital landscape and examining their place in both educational and cultural contexts. The article also outlines prospects for further research and application of internet memes, considering the evolution of digital tools, new platforms, meme dissemination algorithms, and shifting user priorities.

Key words: digital space; internet meme; educational environment; audiovisual culture; verbal elements; sociocultural competence; intercultural communication; communicative-pragmatic potential.

Интернет – это важная составляющая жизни современного общества; коммуникативная среда, создающая совокупность условий, благодаря которым члены социума могут реализовать свое желание или необходимость обмена информацией, используя различные платформы и форматы. Цифровая эпоха предоставляет много возможностей для внедрения новых технологий, которые, опираясь на специфику Интернета (глобальный характер,

выход за государственные и географические границы, открытость, доступность, оперативность в получении и отправке информации, интерактивность), отражают окружающую действительность, продуцируя оригинальные текстовые и визуальные феномены. Одним из таких феноменов является мем – цифровой объект (изображение, текст, образ, идея), «единица интернет-коммуникации, передающаяся от пользователя к пользователю, тиражирующаяся и репродуцирующаяся в интернет-пространстве» [1, с. 122]. Интернет-мем воспринимается как знак современного медиапространства, обладающий определенной культурной и национальной коннотацией и выступающий актуальным средством информативной и аудиовизуальной культуры [2, с. 50–55].

Медиаальный код Интернет-мема способствует параллельному считыванию и восприятию сообщения, поэтому с этих позиций он отвечает важнейшему закону экономии речевых усилий за счет компрессии формы кодирования и выражения смысла.

Мемы, появление которых относится к первым десятилетиям XXI в., классифицировались и анализировались с позиций когнитивной и визуальной социологии, лингвокультурологии, лингвокреативности, герменевтики культуры [3], концепции мультимодальности и прецедентности в массовой культуре [4, с. 122–126], в социокультурном и лингвистическом аспекте [5, с. 79–93].

На фоне существующей исследовательской традиции следует отметить, что в последнее время возросла тенденция изучения и использования мемов в образовательной среде, что способствует активизации внимания, повышению мотивации и развитию когнитивных навыков обучающихся [6, с. 14–15], [7, с. 235–241].

Совокупность вербальной и визуальной составляющих формирует коммуникативно-прагматический потенциал мема, благодаря чему мемы рассматриваются как инструмент межкультурной коммуникации. Являясь катализатором и реакцией на происходящее в мире, мемы часто транслируют мировоззрение людей из разных стран, носителей другого языка и культуры. Мемы способствуют взаимопониманию, создают новые формы культурного обмена и помогают людям выражать мысли и идеи вне зависимости от мировых границ. Поэтому понимание сути и специфики этого явления, его функциональности необходимо для понимания современной реальности и эффективного взаимодействия с пользователями Интернета.

В мемах могут содержаться культурные, исторические, лингвистические отсылки, которые не поддаются переводу [8, с. 8–13]. Для создания определенного коммуникативного эффекта важен контекст, направленный на прецедентную ситуацию или содержащий культурную коннотацию. В подобном декодировании нуждаются распространенные в Интернет-пространстве мультимодальные русскоязычные мемы, в которых изображение сопровождается четверостишием, написанным четырехстопным ямбом, строчными

буквами, без знаков препинания и без рифмы, например: *Изба горела и горела / А конь скакал скакал скакал / Стояли две нерусских бабы / Не зная, что и предпринять* [9, с. 8]. Мем содержит аллюзию на цитату из фрагмента поэмы Н. Некрасова «Мороз Красный нос» («Есть женщины в русских селеньях»): *Коня на скаку остановит/, В горящую избу войдет/*, которая стала языковым стереотипом, характеризующим идеально-героическую русскую женщину. В современной языковой традиции используется для обозначения сильной, здоровой, исключительно русской женщины. В меме же представлены «нерусские бабы», которым не свойственна эта поведенческая характеристика [9, с. 8]. Этот мем актуализирует фоновые знания обучаемых и расширяет их культурный кругозор и коммуникативную компетенцию.

Английский язык в интернет-мемах, являющийся языком межнационального общения, значительно упрощает употребление интернет-мемов пользователями, которые принадлежат к разным культурам и в то же время могут приобщаться к глобальной интернет-культуре, не испытывая языкового барьера. Для понимания мема бывает достаточно минимального уровня владения языком.

Чаще всего в интернет-пространстве существуют одинаковые интернет-мемы на английском и русском языках, и в этом отражается процесс глобализации культурного и информационного контента. Интернет-мемы – важный носитель информации для пользователей социальных сетей: от того, как пользователи из разных стран будут толковать информацию, содержащуюся в одном и том же интернет-меме, зависит успех их общения.

Актуализация чуждых для определенной культуры концептов в интернет-мемах может спровоцировать неадекватное понимание мема или привести к полной коммуникативной неудаче. Чтобы избежать этого, мемы при предъявлении могут комментироваться с учетом подготовки фоновой базы реципиента, или исходя из реалий его культуры.

Адаптация англоязычных мемов наблюдается, например, при переводе с английского языка на другие языки, если в вербальной части мема не актуализируются концепты, принадлежащие к другим культурам. Подобные концепты могут заменяться на привычные, узнаваемые концепты, как это произошло с интернет-мемом про Санта-Клауса, который хочет подарить пользователю Фейсбука словарь на Рождество, намекая на его безграмотность и нарушение языковых норм [10, с. 609]. В аутентичном меме на русском языке социальная сеть *Фейсбук* заменена на соцсеть *ВКонтакте*, что привычнее для русскоязычных пользователей, а *Christmas* (англ. *Рождество*) – на *Новый год* [10, с. 609].

Как это ни парадоксально, но культурный контекст, являющийся неотъемлемой частью мема, не понятый в полной мере представителями других культур, открывает перед «мем-коммуникацией» новые возможности в контексте межкультурного взаимодействия: реципиент, который заинтере-

совался каким-то мемом ввиду его зрелищности, юмористической направленности, популярности, нужности информации и т.д., с большой долей вероятности заинтересуется и источником мема – событием, явлением, процессом, лежащим в основе создания мема. Этот интерес может стать первым шагом к знакомству с новой культурой, а, следовательно, к уменьшению в будущем коммуникативного барьера, барьера непонимания, возникающего между представителями различных культур и к формированию глобального культурного и информационного контента.

Часто непонятны за пределами определенной страны политические или социальные мемы, поэтому для их адекватной интерпретации требуется учитывать современный социокультурный контекст. К примеру, группа «Политические мемы» социальной сети «ВКонтакте» (<https://vk.com/polityumor>) включает мемы, в которых через образы политиков и контекст современных политических событий рассматриваются различные темы: от исторических явлений и культурных достижений до национальной символики. Эти темы выступают как бы негласными этическими ориентирами для пользователей, определяя границы того, над чем можно и нужно смеяться, а над чем следует воздержаться.

При верной расшифровке мема пользователь Интернета интегрируется в новое общество.

Таким образом, в качестве особого коммуникативного знака и культурного объекта нового времени, интернет-мемы влияют на интернет-пользователей, а ускорение передачи информации и неограниченный выбор тем для мемов напрямую влияет как на то место в цифровом пространстве, которое они уже заняли в нем, так и на их будущее развитие.

ЛИТЕРАТУРА

1. Гридина Т. А., Талашманов С. С. Метаязыковой мем: лингвкреативные механизмы порождения и восприятия// Политическая лингвистика. 2020. № 2 (80). С. 134–143.
2. Гузаерова Р. Р. Интернет-мем как знак современного медиапространства // Филология и культура. 2017. № 2 (48). С. 50–55.
3. Олешкова А. М. Интернет-мем как способ воспроизводства культуры // Общество: философия, история, культура. 2017. № 8. DOI: <https://doi.org/10.24158/fik.2017.8.36>.
4. Канашина С. В. Интернет-мем и прецедентный феномен // Вестник Томского государственного педагогического университета. 2018. № 4 (193). С. 122–126.
5. Блинова О. А. Мультиmodalность в сетевом политическом дискурсе: интернет-мемы о независимости Шотландии // Научный диалог. 2019. № 10. С. 79–93.

6. Загоруйко А. О., Ефремова М. А. Потенциал использования Интернет-мемов в качестве обучающего средства // Вопросы методики преподавания в вузе. 2019. № 28. С. 14–15.
7. Резцова С. А. Лингводидактический потенциал интернет-мема как мультимодального текста // Язык в эпоху цифровых трансформаций и развития искусственного интеллекта. Минск: МГЛУ, 2024. С. 235–241.
8. Назаренко А. Л. К вопросу о цифровизации (лингвистического) образования // Актуальные вопросы цифровизации лингвистического образования в вузе. М., 2022. С. 8–13.
9. Керимова С. А. Мемы как новое явление в Интернет-пространстве: социокультурный и лингвистический аспекты // Русский язык и литература в Азербайджане. 2020. № 1–2. С. 4–9.
10. Канашина С.В. Интернет-мем в контексте межкультурной коммуникации // Магия ИННО: новые измерения в лингвистике и лингводидактике: сб. науч. трудов, В 2 Т., Т. 1. М: МГИМО–Университет, 2017. С. 606–612.

УДК 81'33

Кустова Ольга Юрьевна

кандидат филологических наук,
доцент
ФГБОУ ВО Российский государст-
венный педагогический университет
им. А. И. Герцена
г. Санкт-Петербург, Россия

Olga Kustova

PhD in Philology, Associate Professor
Herzen State Pedagogical University
of Russia
Saint-Petersburg, Russia
o_kustova@mail.ru

Шарешик Денис Юрьевич

магистрант
ФГБОУ ВО Российский государст-
венный педагогический университет
им. А. И. Герцена
г. Санкт-Петербург, Россия

Denis Shareshik

Master's Student
Herzen State Pedagogical University
of Russia
Saint-Petersburg, Russia
shareshik.denis@yandex.ru

ЛИНГВОКУЛЬТУРНАЯ СПЕЦИФИКА ПРОЯВЛЕНИЯ ЭМОЦИЙ В ПОЛИКОДОВЫХ ТЕКСТАХ В КОНТЕКСТЕ ЦИФРОВОЙ КОММУНИКАЦИИ

Статья посвящена вербальным и невербальным средствам репрезентации позитивных эмоций в поликодовых текстах русской и финской лингвокультур. Лингвокультурологический и семиотический подходы к исследованию текстов графических романов позволили выявить их ключевые различия: тексты на русском языке характеризуются экспрессивностью, избыточностью сигналов и яркой визуализацией, на финском языке – сдержанностью, лаконичностью и минимализмом. Методика исследования, представляющая собой синтез доминантного, компонентного, лингвокультурологического и контрастивного анализа, предусматривала определение функций и адресата текста, анализ стиливых черт поликодового текста в этнокультурном плане, описание языковых признаков на основе этнокультурной информации, выявление культурных связей и их стилистических возможностей. Результаты исследования демонстрируют влияние культурных кодов на эмоциональную концептосферу и подтверждают гипотезу о национальной специфике эмотивной коммуникации. Особое внимание уделяется трансформации средств экспрессии в цифровой коммуникации. Исследование носит прикладной характер и является ключевым этапом изучения вербализации и визуализации эмоций в чат-бот коммуникации.

Ключевые слова: поликодовый текст; позитивные эмоции; русская лингвокультура; финская лингвокультура; мультимодальность; цифровая коммуникация.

LINGUOCULTURAL SPECIFICITY OF THE MANIFESTATION OF EMOTIONS IN POLYCODE TEXTS IN THE CONTEXT OF DIGITAL COMMUNICATION

The article is devoted to identifying the specifics of verbal and non-verbal means of positive emotions representation in polycode texts of Russian and Finnish linguocultures. Linguocultural and semiotic approaches to the study of texts of graphic novels have allowed us to identify their key differences: texts in Russian are characterized by expressiveness,

redundancy of signals and vivid visualization, while texts in the Finnish language – by restraint, brevity and minimalism. The research methodology, which is a synthesis of dominant, component, linguocultural and contrastive analysis, included the definition of the functions and addressee of the text; the analysis of style features of the polycode text in ethnocultural terms; description of linguistic features based on ethnocultural information; identification of linguistic features of the text in ethnocultural terms. The results of the study demonstrate the influence of cultural codes on the emotional conceptosphere and confirm the hypothesis about the national specificity of emotive communication. Special attention is paid to the transformation of means of expression in the digital communication. The research is of applied nature and is a key stage in the study of verbalization and visualization of emotions in chatbot communication.

Key words: polycode text; positive emotions; Russian linguoculture; Finnish linguoculture; multimodality; digital communication.

Современная лингвистика характеризуется возрастающим интересом к изучению взаимосвязи языка, культуры и эмоций. В условиях глобализации и развития цифровых технологий особую актуальность приобретает исследование способов выражения эмоций в поликодовых текстах, которые объединяют вербальные и невербальные компоненты для передачи различных типов информации. Лингвокультурологический анализ проявления позитивных эмоций в текстах разных типов позволяет выявить национально-специфические особенности эмоциональной концептосферы различных языковых сообществ.

Поликодовые тексты представляют собой сложные семиотические образования, в которых «сообщение закодировано семиотически разнородными средствами – вербальным и невербальным компонентами, объединение которых представляет собой определенную структуру, характеризующуюся проявлением взаимозависимости составляющих как в содержательном, так и в формальном аспектах» [1, с. 72]. В таких текстах невербальные знаки являются носителями информации или вносят дополнительные оттенки в содержание, что делает их особенно значимыми для передачи эмоциональных состояний.

Лингвокультурологический анализ поликодовых текстов требует применения комплексного подхода, основанного на принципе междисциплинарности. Основными методами настоящего исследования послужили: доминантный анализ для выделения наиболее частотных лексем, обозначающих эмоциональные состояния; компонентный анализ для выявления национально-культурного компонента значения эмотивных единиц; анализ лингвокультурологического поля как совокупности языковых единиц, объединенных общим культурным смыслом; контрастивный анализ для сопоставления способов выражения эмоций в русской и финской лингвокультурах.

Материал исследования составили два поликодовых художественных произведения: графический роман «Tumpeät tytöt» Риины Тансканен [2] на финском языке и графический роман Аскольда Акишина и Кирилла Кутузова «Омон Ра» [3] по мотивам одноимённого произведения Виктора Пелевина на русском языке. В эмпирический корпус вошли по 20 фрагментов из каждого

романа, в которых как на вербальном, так и на невербальном уровнях наиболее ярко прослеживаются эмоции персонажей и характерные для произведений культуроспецифичные особенности их проявления.

Теоретические предпосылки исследования

Основанием настоящего исследования является лингвокультурологический подход к исследованию эмоций, базирующийся на концепции, согласно которой культура и язык представляют собой формы сознания, отражающие человеческую деятельность [4, с. 105]. Как известно, существуют сложные взаимосвязи между культурой, языком и эмоциями, которые на вербальном уровне проявляются в семантике слов и выражений, а также находят отражение в идиомах, поговорках и конвенциональных метафорах.

В лингвистике эмоций выделяются два ключевых термина: «лексика эмоций», которая ориентирована на объективацию психических состояний в языке, и «эмоциональная лексика», выполняющая экспрессивную и прагматическую функции. Эмоции проявляются в языке двояко: во-первых, как эмоциональное сопровождение, возникающее в результате выражения эмоционального состояния говорящего; во-вторых, как отражение объективно существующей реальности через языковые знаки [5, с. 12–14]. Современная лингвистика эмоций фокусируется на изучении эмотивности в коммуникативно-функциональном и прагматическом аспектах. Актуальными направлениями являются изучение эмотивных параметров интернет-коммуникации и межкультурной коммуникации в зеркале эмоционального взаимодействия.

Поликодовый текст, являющийся предметом данного исследования, рассматривается как структурно-содержательное единство, организованное в комбинации естественного языка с элементами других знаковых систем или упорядоченных множеств. Для обозначения таких текстов применяются различные термины, такие как «креолизованные тексты», «полисемические тексты», «синкретические тексты», «лингвовизуальные», «изовербальные комплексы». Тем не менее термин «поликодовый» представляется наиболее релевантным для характеристики материала данного исследования. Поликодовые тексты характеризуются следующими особенностями: одновременное использование различных семиотических кодов (вербального и невербального); взаимодействие семиотически разнородных элементов; создание определенной структуры с взаимозависимостью составляющих в содержательном и формальном аспектах.

В современной лингвистике наблюдается тенденция к использованию термина «мультимодальный текст» для обозначения цифровых текстов [6, с. 45]. Мультимодальные тексты определяются как тексты, в которых «коммуникативная ситуация представлена с использованием нескольких модальностей восприятия, конкретно зрения и слуха» [7, р. 145], при этом базовое распределение смысловых компонентов на лингвистические (вербальные) и экстралингвистические (невербальные), принятое для исследования поликодовых текстов, сохраняет свою актуальность. Исследование мультимодальных текстов приобретает особое значение в условиях цифровизации, которые

становятся доминирующими в современной коммуникации [8], [9]. Развитие технологий приводит к появлению новых форм выражения эмоций, включая визуальные средства передачи эмоций в цифровом пространстве.

Культуро-специфичные особенности выражения позитивных эмоций

Русская лингвокультура характеризуется богатством средств выражения эмоциональных состояний, исследователи сходятся во мнении, что русский язык предлагает больше средств для выражения эмоций по сравнению с другими языками. В целом для русской коммуникативной культуры характерна эмоциональность, которая проявляется в принципе «говорю, что чувствую» [10, с. 36–45]. Так, на лексическом уровне эмоция радости выражается в русском языке следующими средствами: эмоционально окрашенными аффективными прилагательными с положительной коннотацией, фразами-клише, междометиями, а также эмотивами, которые произошли от вокативных форм имен существительных (*батюшки, матушки, мамочки*) [11]. Русские номинации эмоций сочетаются с прилагательными, описывающими различные черты характера: *болтливая радость, ревнивый гнев*, что свидетельствует о детализированном восприятии эмоциональных состояний в русской культуре.

Финская культура, напротив, демонстрирует более сдержанный подход к выражению эмоций. Исследования коммуникативного поведения финнов показывают, что они отличаются эмоциональной сдержанностью и склонностью к подавлению эмоций. В финской культуре воспитание и образование сдерживают избыточную мимику и спонтанное выражение радости, любви, грусти и других чувств [12, р. 274]. Финский язык располагает следующими средствами выражения позитивных эмоций: базовые номинации: *ilo* '*радость*', *onnellinen* '*счастливый*'; эмоциональные фразы удивления и одобрения: *Että pitikin!* '*Надо же!*', *Uskomatonta!* '*Невероятно!*'; устойчивые выражения для передачи радости: *Totta kai* '*Точно*', *Tietenkin!* '*Конечно!*'. Важной особенностью финского коммуникативного поведения является принцип «говори, когда уверен». Финны демонстрируют сдержанность в выражении эмоций и предпочитают нейтральную лексику даже в ситуациях эмоционального напряжения.

Понимание специфики проявления эмоций в русской и финской лингвокультурах важно при разработке более точных и адаптивных алгоритмов обработки естественного языка. Учет культурных различий в выражении эмоций позволяет создавать модели, которые учитывают контекст и культурные особенности, что особенно важно для приложений, связанных с анализом настроений, чат-ботами и виртуальными помощниками. Эти данные релевантны при создании более эффективных систем, способных адекватно интерпретировать и генерировать текст в зависимости от культурного контекста.

Проявление позитивных эмоций в поликодовых текстах

В поликодовых текстах вербальные средства выражения позитивных эмоций включают различные уровни языковой системы и комплекс невер-

бальных элементов. На лексическом уровне используются: эмотивная лексика с положительной коннотацией, междометия и эмоциональные восклицания, фразеологические единицы, отражающие культурные коды радости и удовольствия. На синтаксическом уровне эмоциональность передается через восклицательные конструкции, повторы и эмфатические структуры, специфические интонационные модели. Невербальная составляющая поликодовых текстов включает, в свою очередь, визуальные элементы (изображения, схемы, цветовое оформление), типографические средства (размер, цвет, стиль шрифта), композиционные решения (расположение элементов, использование пространства).

К культуроспецифическим особенностям поликодовых текстов на русском языке относят выраженность эмоциональности вербального компонента, использование ярких визуальных элементов, тенденцию к избыточности в передаче позитивных эмоций, активное использование междометий и восклицаний [11]. В то же время поликодовые художественные тексты на финском языке отличаются сдержанностью в выражении эмоций, предпочтением нейтральных формулировок, минималистичным подходом к визуальному оформлению, точностью и лаконичностью эмоциональных проявлений.

Анализ вербальных и невербальных средств проявления эмоций имеет большое значение для развития технологий искусственного интеллекта. Вербальные средства, такие как выбор слов, структура предложений, интонация, могут быть использованы для обучения моделей, которые распознают и интерпретируют эмоциональные состояния пользователей. Невербальные средства, такие как жесты, мимика, могут быть интегрированы в системы распознавания эмоций, что позволит сделать коммуникацию в цифровой среде более полноценной.

Результаты исследования

Лингвокультурологический анализ проявления позитивных эмоций в поликодовых текстах на русском и финском языках выявляет существенные различия в способах концептуализации и вербализации эмоциональных состояний. В графическом романе на русском языке «Омон Ра» положительные эмоции отображаются нарочито подчёркнуто: персонажи открыто выражают свои эмоции вербально, используя экспрессивные лексические единицы и подкрепляя их широкой улыбкой. Финский графический роман, в свою очередь, характеризуется наибольшей сдержанностью: персонажи чаще всего неявно выражают радость и другие положительные эмоции, в ряде случаев положительные эмоции персонажей прослеживаются лишь через анализ взаимосвязи вербальной и невербальной составляющей поликодового текста. Таким образом, русская лингвокультура демонстрирует большую эмоциональную экспрессивность и разнообразие средств выражения позитивных эмоций, в то время как финская культура характеризуется минимализмом в эмоциональных проявлениях. Поликодовые художественные тексты представляют собой особую сферу функционирования эмоциональной лексики, где синергия вербальных и невербальных компонентов создает новые возможности для передачи культуро-специфичных смыслов.

Развитие цифровых технологий и мультимодальной коммуникации расширяет арсенал средств выражения эмоций и требует разработки новых методологических подходов к их изучению. Дальнейшие исследования в данной области должны учитывать динамику развития цифровой коммуникации и формирование новых форм эмоциональной выразительности в мультикультурном пространстве. Особый интерес представляет изучение механизмов адаптации традиционных способов выражения эмоций к современным формам поликодовой коммуникации, разработка мультимодальных лингвистических корпусов и использование данных о специфике вербального и невербального проявления эмоций в машинном обучении. В контексте развития технологий искусственного интеллекта результаты исследования применимы при создании адаптивных интерфейсов, разработке систем поддержки, для улучшения качества общения в цифровой среде. Чат-боты и виртуальные помощники, обученные на основе культурных особенностей коммуникации, смогут более точно реагировать на запросы пользователей, учитывая их эмоциональное состояние.

Финансирование: Исследование выполнено за счет внутреннего гранта РГПУ им. А. И. Герцена (проект № 75-ВГ).

ЛИТЕРАТУРА

1. Анисимова Е. Е. Паралингвистика и текст (к проблеме креолизованных и гибридных текстов) // Вопросы языкознания. 1992. № 1. С. 71–78.
2. Tanskanen R. *Tympeät tytöt*. Helsinki : Into Kustannus, 2021. 160 p.
3. Пелевин В. О., Акишин А., Кутузов К. Омон Ра. Москва : Комильфо. Комиксы русских авторов, 2018. 145 с.
4. Тарасов Е. Ф. Язык и культура: методологические проблемы // Язык. Культура. Этнос. Москва : Наука, 1994. С. 105–113.
5. Бабенко Л. Г. Лексические средства обозначения эмоций в русском языке. Свердловск : Изд-во Уральского университета, 1989. 211 с.
6. Некрасова Е. Д. К вопросу о восприятии полимодальных текстов // Вестник Томского государственного университета. 2014. № 378. С. 45–48.
7. Сидорова Е. В. Использование мультимодального текста при изучении эмоциональной речи // Ученые записки ОГУ. Серия: Гуманитарные и социальные науки. 2013. №5. С. 145–149.
8. Кожевникова Л. П. Мультимодальный дискурс в цифровой среде // Вопросы языкознания. 2023. № 1. С. 74–87.
9. Кустова О. Ю., Полтавец Г. А. Передача национально-прецедентных интертекстуальных компонентов при локализации видеоигр [Электронный ресурс] // Вестник Пермского национального исследовательского политехнического университета. Проблемы языкознания и педагогики. 2024. № 1. С. 8–18. DOI <https://doi.org/10.15593/2224-9389/2024.1.1>

10. Ларина Т. В. Категория вежливости и стиль коммуникации: Сопоставление английских и русских лингвокультурных традиций. Москва : Рукописные памятники Древней Руси, 2009. 512 с.
11. Абросимова Л. С., Ключко Н. Г. Языковая репрезентация эмоции «радость» в русской и английской современной прозе // Вестник ВГУ. Сер. Лингвистика и межкультурная коммуникация. 2020. № 3. С. 63–68.
12. Sajavaara K, Lehtonen J. The silent Finn revisited // Jaworski, Adam (ed.), Silence: Interdisciplinary Perspectives. Berlin: Mouton de Gruyter, 1997. P. 263–283.

УДК 81'42 + 811.929.1

Марченко Татьяна Владимировна
кандидат филологических наук,
доцент
Северо-Кавказский федеральный
университет
г. Ставрополь, Россия

Tatiana Marchenko
PhD in Philology, Associate professor
North-Caucasus Federal University
Stavropol, Russia
tatiana-marchenko-25@yandex.ru

ПОЛИКОДОВАЯ ДИСКУРСИВНАЯ РЕПРЕЗЕНТАЦИЯ ДИФФЕРЕНЦИАЛЬНЫХ ПРИЗНАКОВ И АТТРИБУТОВ ПРЕЦЕДЕНТНЫХ ФЕНОМЕНОВ В МЕДИАПРОСТРАНСТВЕ

Цель статьи состоит в анализе специфики семиотической репрезентации дифференциальных признаков прецедентных феноменов в поликодовых текстах. Материалом для анализа послужили семиотически гетерогенные тексты в жанре комического – Интернет-мемы, манифестирующие прецедентные референции в одном из задействованных семиотических кодов. Взаимоотношение ядерных и периферийных компонентов значения и атрибутов в прецедентном феномене носит динамичный характер, опосредованный перманентным влиянием дискурсивной практики и актуального экстралингвистического контекста. В условиях семиотической вариативности референциальных элементов дискурсивная репрезентация дифференциальных признаков прецедентного феномена в поликодовых текстах манифестирует конвергентный характер. Сочетаемость репрезентантов определяется семиотической природой самого прецедентного феномена (вербальный, визуальный, аудиальный или комплексный) и спецификой текста (динамичный / статичный характер). Смыслопорождение в малых текстовых единицах, содержащих репрезентанты прецедентных феноменов, опирается на механизмы когнитивного брендинга, согласующие в соответствии с интенцией автора разные концептуальные пространства в единое целое.

Ключевые слова: прецедентный феномен; визуальный прецедентный феномен; поликодовый текст; интернет-мем; медиaproстранство.

POLYCODE DISCOURSE REPRESENTATION OF DIFFERENTIAL FEATURES AND ATTRIBUTES OF PRECEDENT PHENOMENA IN MEDIASPACE

The article analyzes peculiarities of semiotic representation of differential features of precedent phenomena in polycode texts. The material for the analysis is semiotically heterogeneous texts in the comic genre – Internet memes, manifesting precedent references in one of the semiotic codes involved. The relationship between the core and peripheral components of meaning and attributes in the precedent phenomenon is dynamic, preconditioned by the permanent influence of discursive practice and the current extralinguistic context. In the conditions of semiotic variability of referential elements, discursive representation of differential features of the precedent phenomenon in polycode texts manifests convergent nature. The compatibility of representatives is determined by the semiotic nature of the precedent phenomenon itself (verbal, visual, auditory or complex) and the specificity of the text (dynamic /

static nature). Meaning generation in small text units containing representatives of precedent phenomena relies on cognitive blending mechanisms, which coordinate different conceptual spaces into a single whole in accordance with the author's intention.

Key words: precedent phenomenon; visual precedent phenomenon; polycode text; Internet meme; media space.

Многоканальность современной медийной коммуникации, опирающаяся на широкий спектр технических и технологических ресурсов, конвергенция различных дискурсивных пространств и стирание четких жанрово-стилистических границ особым образом опосредуют специфику инкодирования всех видов информации [1], в частности прецедентной. Прецедентные феномены, служащие особыми гносеологически и аксиологически релевантными культуроспецифичными знаками, маркируют различные ипостаси культурной идентичности (например, коллективной / индивидуальной, общецивилизационной / национальной / локальной) и выступают в качестве средств с амбивалентным функциональным потенциалом: эстетическим, манипулятивным, людическим, парольным, оценочным и др. Данная специфика детерминирует повсеместное использование прецедентных феноменов как в институциональных, так и неинституциональных типах дискурсивных пространств [2]. Распространение семиотически гетерогенных текстов, в свою очередь, актуализирует необходимость научного осмысления семиотически вариативной дискурсивной объективации прецедентных феноменов разных типов [3], нарративизации формирования прецедентных единиц [4], закрепления монокодовых и поликодовых прецедентных единиц [5], а также прецедентных единиц конвергентной вербально-аудиально-визуальной семиотической природы [6].

Цель данной статьи состоит в анализе специфики семиотической репрезентации дифференциальных признаков прецедентных феноменов в поликодовых текстах. Материалом для анализа послужили полисемиотические тексты в жанре комического – Интернет-мемы, манифестирующие прецедентные референции в одном из задействованных семиотических кодов.

Прецедентный феномен определяется В. В. Красных как ядерный элемент когнитивной базы носителей определенного ментально-лингвального комплекса, соответствующий критериям известности, рекуррентности, эмоциональной и познавательной релевантности [7, с. 170]. Каждый прецедентный феномен обладает некоторой совокупностью дифференциальных признаков, детерминирующих его семантику и формирующих сложную структуру ядерных и периферийных значений. Атрибуты, в свою очередь, являются ассоциируемыми, но опциональными элементами-сигнификатами.

Взаимоотношение ядерных / периферийных компонентов значения и атрибутов носит динамичный характер, находящийся под перманентным влиянием дискурсивной практики и актуального экстралингвистического

контекста. В условиях семиотической вариативности референциальных элементов дискурсивная репрезентация дифференциальных признаков прецедентного феномена в поликодовых текстах носит конвергентный характер:

- дифференциальные признаки и / или атрибуты одного прецедентного феномена актуализируются разными семиотическими кодами;
- дифференциальные признаки и / или атрибуты разных прецедентных феноменов актуализируются разными семиотическими кодами.

Сочетаемость репрезентантов определяется семиотической природой самого прецедентного феномена (вербальный, визуальный, аудиальный или комплексный) и спецификой текста (динамичный / статичный характер).

Проиллюстрируем наши наблюдения примерами из картотеки нашего исследования. Так, Интернет-мем [9] является поликодовым текстом, инкорпорирующим вербальный и визуальный коды. Вербальный код представлен двумя фрагментами: заголовочным комплексом и резюме. Первый фрагмент интродуцирует ситуацию: *В социальной сети фотография лошади на поле, вокруг туман. Подпись на английском про символику лошади. Что-то такое очень философское и пафосное. Куча комментариев на английском на тему: как красиво, как умно, как глубоко и т.п.* Второй фрагмент – краткий вывод, объясняющий суть сообщения и ставящий под сомнение целесообразность действий, обозначенных в первом фрагменте: *И одинокий коммент на русском: Ееееежишиик!*

Визуальный код сообщения представляет фотографическое изображение большого открытого пространства с едва уловимыми очертаниями пейзажа в тумане. Наиболее отчетливо виднеется лишь один объект – вороная лошадь, морда которой направлена анфас.

Поликодовый текст содержит референцию к советскому мультипликационному фильму «Ежик в тумане» (1975), основная сюжетная линия которого посвящена путешествию главного героя – Ежика – в гости к другу Медвежонку. Мир, погруженный в предвечерний туман, транслирует удивительные образы, пугающие и манящие главного героя. Одним из таких образов является белая Лошадь, силуэт которой неоднократно появляется и исчезает в дымке. Несколько раз Ежик и Медвежонок протяжно окликают друг друга. Итогом путешествия является воссоединение друзей и последующее чаепитие.

Артефакт анимационного творчества широко известен представителям русскоязычного лингвокультурного сообщества и неоднократно становился предметом анализа в рамках разных предметных областей: теория кино, психология, культурология и др. Исследователи и рядовые реципиенты склонны усматривать в кинопроизведении аллегорию жизненного пути, возникающих на этом пути сложностей и искушений, поисков себя и своего места в мире и многое другое.

Дифференциальные признаки прецедентного феномена семиотически вариативны: непосредственно визуальный образ Ежика в тумане в формате кадра и мультипликационного фильма или ином варианте изображения; вербальные репрезентанты – название самого мультфильма и реплики главных персонажей: *Ежик!; Медвежонок!; Вы для меня самые лучшие из тех, кто есть на земле!; Если тебя нет, то и меня нет; Я сегодня проснулся и понял, что лучше тебя нет на свете; И еще Ежик думал о Лошади, как она там в тумане* и др.

Атрибутами прецедентного феномена также служат семиотически гетерогенные элементы, к числу которых представляется возможным отнести фрагменты визуального ряда, образы других персонажей, отдельные предметы (например, котомка главного героя), некоторые особенности колористического и визуального оформления (цветовая гамма и эстетика графического стиля). Вербальные атрибуты незначительны, так как кинопроизведение отличается лаконичностью и содержит минимальное количество реплик. Паравербальные атрибуты включают специфическую просодику речи героев.

Интернет-мем содержит вербальный дифференциальный признак анализируемого прецедентного феномена: протяжную реплику-восклицание *Ееееежиииик!* В визуальном коде актуализируются отдельные атрибуты прецедентного феномена, воспроизведение которых носит реминисцентный характер, не отличающийся изоморфизмом. Так, образ лошади представлен фотографическим изображением без дополнительной художественной обработки и в ином цветовом исполнении – черном.

С точки зрения смыслопорождения, установление корреляции между обозначенными вербализованным дифференциальным признаком и визуализированными атрибутами обеспечивает референцию к прецедентному феномену и выступает примарным условием для последующего когнитивного блендинга, результатом которого выступает ирония. Основу иронии составляют два элемента. Первый – поликодовая референция к прецедентному феномену во всей совокупности его значений, делимитация которых не обеспечивается контекстом сообщения. Второй – текст в заголовке, обобщающий нарочито лишенную фоновых знаний обывательскую интерпретацию визуального образа в контексте внешнего лаконизма и внутреннего глубокого посыла. Механизм иронии опирается на противопоставление сложности и простоты, исключительности и обыденности, общекультурного и культуроспецифического, непрецедентного и прецедентного.

Таким образом, дискурсивная репрезентация дифференциальных признаков и атрибутов прецедентных феноменов в медиапространстве носит поликодовый характер. Смыслопорождение в малых текстовых единицах, содержащих репрезентанты прецедентных феноменов, опирается на механизмы когнитивного блендинга, согласующие в соответствии с интенцией автора разные концептуальные пространства в единое целое.

ЛИТЕРАТУРА

10. Бредихин С. Н., Каменский М. В., Шибкова О. С. Новые медиа в процессе модификации концептуально-валерного пространства // Вестник Воронежского государственного университета. Серия: Филология. Журналистика. 2024. № 2. С. 107–111.
11. Мардиева Л. А. Визуальная прецедентность как объект лингвистических исследований: современное состояние и перспективы развития // Ученые записки Казанского университета. Сер. Гуманитарные науки. 2023. № 3 (165). С. 19–32.
12. Марченко Т. В. Маркеры поликодовой объективации прецедентных феноменов в семиотически осложненном тексте // Гуманитарные и юридические исследования. 2024. № 2 (11). С. 388–395.
13. Марченко Т. В., Бредихин С. Н., Лепилкина О. И. Специфика формирования ассоциативного потенциала нового прецедентного феномена и поликодовой объективации в современном британском медиадискурсе // Вестник Калмыцкого университета. 2023. № 3 (59). С. 109–115.
14. Марченко Т. В. Особенности семиотического инкодирования прецедентного имени или «простой советский человек» в поликодовом тексте // Гуманитарные и юридические исследования. 2024. № 1 (11). С. 175–190.
15. Марченко Т. В. Нарративная природа формирования и функционирования комплексных прецедентных феноменов // Научная мысль Кавказа. 2023. № 3 (115). С. 135–140.
16. Красных В. В. «Свой» среди «чужих»: миф или реальность? Москва : Гнозис, 2003. 375 с.
17. Бредихин С. Н. Теория некооперативных игр в порождении и интерпретации высказываний институционального дискурса // Профессиональная коммуникация: актуальные вопросы лингвистики и методики. 2023. № 16. С. 50–58.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

18. И одинокий коммент на русском URL: https://s00.yaplakal.com/pics/pics_original/1/0/5/17921501.jpg/ (дата обращения: 03.07.2025).

УДК 81-11

Обухова Надежда Владимировна

кандидат педагогических наук, доцент
кафедра романо-германской
филологии
Университет Мамуна
г. Хива, Узбекистан

Nadejda Obuxova

PhD in Pedagogy, Associate Professor
Department of Romano-Germanic
Philology
Mamun University
Khiva, Uzbekistan
tabithanvo@gmail.com

МЕЖКУЛЬТУРНАЯ МЕДИАКОММУНИКАЦИЯ В ЦИФРОВУЮ ЭПОХУ: СОПОСТАВЛЕНИЕ АНГЛОЯЗЫЧНЫХ, РУССКОЯЗЫЧНЫХ И УЗБЕКОЯЗЫЧНЫХ МЕДИЙНЫХ ФРЕЙМОВ

Статья посвящена фреймовому анализу концепта *cancel culture* в англоязычном, русскоязычном и узбекоязычном медиадискурсе в условиях цифровой трансформации. На материале актуальных медиатекстов рассматриваются особенности формирования и трансформации фреймов «отмены» в различных культурных и политических контекстах. Показано, что в англоязычном пространстве *cancel culture* репрезентируется как моральная санкция, социальный механизм давления или ироничный медийный шаблон. В русскоязычных СМИ термин приобретает идеологически окрашенное значение, включая интерпретацию как угрозу национальной культуре. В узбекоязычном дискурсе отмечается отсутствие устойчивого фрейма *cancel culture*, что интерпретируется как фреймовая лакуна; однако выявлены функциональные аналоги – практики институционального исключения. Исследование основано на принципах медиалингвистики и фрейм-анализа, демонстрируя потенциал межкультурной интерпретации смыслов в условиях цифровой медиасреды.

Ключевые слова: cancel culture; фреймовый анализ; медиадискурс; цифровая среда; межкультурная коммуникация; культурная лакуна.

INTERCULTURAL MEDIACOMMUNICATION IN DIGITAL ERA: THE COMPARATIVE ANALYSIS OF ENGLISH, RUSSIAN AND UZBEK MEDIAFRAMES

This article presents a frame-based analysis of the concept of *cancel culture* as represented in English-, Russian-, and Uzbek-language media discourse in the context of digital transformation. Using current media texts, the study explores the formation and transformation of «*cancellation*» frames across different cultural and political settings. The analysis shows that in English-language discourse, *cancel culture* appears as a mechanism of moral sanction, social pressure, or an ironic media construct. In Russian-language media, the term acquires an ideologically charged meaning, including interpretations framing it as a threat to national identity. In Uzbek-language discourse, a stable frame of *cancel culture* is absent, which is interpreted as a frame lacuna; nevertheless, functional analogues – such as institutional exclusion – are identified. Grounded in media linguistics and frame theory, the study demonstrates the interpretive potential of cross-cultural meaning analysis within the digital media environment.

Key words: cancel culture; frame analysis; media discourse; digital environment; cross-cultural communication; cultural lacuna.

Современные тенденции развития материально-технической базы информационно-коммуникационных систем характеризуются бурным распространением цифровых технологий, появлением принципиально новых средств коммуникаций на фоне не виданной прежде гиперкоммерциализации процессов массовой коммуникации. Как показывают исследования ценностных ориентаций, вкусов и потребностей населения, цифровые технологии не только удовлетворяют новым требованиям динамично живущих людей, но и кардинальным образом меняют их отношение к деятельности медиасистемы. Анализ тенденций развития медиаиндустрии показывает, что за десять лет доля интернета на мировом рынке выросла в четыре раза, в то время как доля печатной прессы сократилась в три раза. Бурными темпами растет медиапотребление сетевых СМИ, новостных агрегаторов и социальных сетей [1, с. 8].

Медиапространство, в котором на сегодняшний день сосуществуют все формы контента и коммуникации – от новостных сводок до пользовательских мемов, – не только отражает действительность, но и активно формирует её, моделируя культурные и социальные нормы, регулируя способы интерпретации информации. В связи с этим особую актуальность приобретает подход Т. ван Дейка, согласно которому современные медиа обладают дискурсивной властью, позволяющей осуществлять контроль над общественным сознанием посредством повторяемых структур нарратива, тем и ролей [2]. Такая власть осуществляется не через прямое принуждение, а через фреймирование – когнитивную организацию информации, которая направляет восприятие и интерпретацию событий.

В условиях межкультурной коммуникации фреймы, встроенные в медиатексты, могут не совпадать с когнитивными схемами носителей других языков и культур. Это особенно актуально при взаимодействии медиадискурсов, обладающих различной ценностной базой, коммуникативной нормой и культурной моделью интерпретации. Так, лексически понятные конструкции могут не вызывать аналогичных фреймовых ассоциаций, что ведёт к частичному или полному искажению смысла.

В таком контексте межкультурный фреймовый анализ становится эффективным инструментом для выявления расхождений не только на уровне языка, но и на уровне когнитивной картины мира, что особенно важно при изучении медиатекстов в поле межкультурного взаимодействия. Выбор для сопоставительного анализа англоязычных, русскоязычных и узбекоязычных медиатекстов обусловлен, с одной стороны, актуальностью этих языков в глобальном и региональном медиапространстве, а с другой – возможностью прямого наблюдения и погружения в узбекоязычную медиасреду благодаря личному пребыванию автора в Узбекистане. Это позволяет не только изучать языковой материал в его актуальном употреблении, но и учитывать контекстуальные, культурные и прагматические особенности, влияющие на интерпретацию медийных фреймов. При сопоставлении одного и того же культур-

ного концепта или информационного повода могут активироваться разные фреймы, вызывать различные оценки, и, как следствие, разные модели поведенческого восприятия. Таким образом, проблема фреймовой несовместимости приобретает как лингвистическую, так и социокультурную значимость.

Целью настоящего исследования является выявление особенностей фреймовой репрезентации медийных смыслов в англоязычных, русскоязычных и узбекоязычных медиатекстах, а также анализ фреймовых трансформаций и возможных искажений, возникающих при их интерпретации в условиях межкультурной коммуникации и цифровой медиасреды. В задачи исследования входят следующие шаги:

- 1) Описать теоретические основы понятия фрейма и медиафрейма в контексте когнитивной и медиалингвистики;
- 2) Обосновать значимость фреймового подхода для анализа межкультурной медиакommunikации в условиях цифровой трансформации;
- 3) Выявить культурно-специфичные фреймы в англоязычных, русскоязычных и узбекоязычных медиатекстах;
- 4) Проанализировать случаи фреймовых расхождений и искажений при интерпретации медийных сообщений в сопоставлении трёх культурных контекстов.

Объектом исследования является межкультурная медиакommunikация в условиях цифровой среды. Под предметом исследования мы понимаем фреймовые особенности репрезентации и интерпретации медийных смыслов в англоязычных, русскоязычных и узбекоязычных медиатекстах.

Методы исследования, применявшиеся в работе, включают качественный фреймовый анализ, сопоставительный анализ медиатекстов, элементы дискурсивного анализа (в частности, подход Т. ван Дейка), контекстуальный и интерпретационный анализ с привлечением когнитивной и культурно-ориентированной лингвистики.

В качестве материала исследования были отобраны англоязычные, русскоязычные и узбекоязычные медиатексты, опубликованные на различных цифровых платформах (онлайн-издания, блоги, новостные сайты, социальные медиа) в период с 2021 по 2024 гг. Отбор текстов осуществлялся на основе тематической релевантности: преимущественно анализировались публикации, затрагивающие социально-политические темы, вызывающие разную фреймовую интерпретацию в разных медиакультурах.

При выборе материала учитывалась возможность сопоставления одного и того же концепта или темы в разных культурных и языковых средах. В частности, были использованы тексты из таких источников, как The Telegraph, информационного агентства Reuters, the Atlantic, The Times, Москва 24, Газета.ру, Ozodlik Radio, Asia Tass, Gazeta.uz и др. Также в анализ включены примеры пользовательских высказываний и метафор, распространённых в цифровом дискурсе, в том числе в Telegram-каналах и социальных сетях.

Ввиду ограничения объёма исследования, выборка текстов носила целевой характер и была ориентирована не на количественный охват, а на глубинную интерпретацию фреймовой структуры и сопоставление межкультурных различий.

Обсуждения и результаты

Анализ современных исследований в области медиадискурса демонстрирует устойчивую направленность научного интереса к вопросам, обусловленным актуальными политическими, социальными и экономическими реалиями глобального сообщества. Интенсификация межкультурного взаимодействия на фоне стремительных геополитических трансформаций определяет ключевые каналы коммуникации между культурами – прежде всего средства массовой информации, а в условиях цифровизации – цифровые медиаплатформы.

В ряде научных работ (Шорова, 2024, Денисова, 2023) [3, 4] подчёркивается возрастающая роль межкультурной интерпретации в цифровом медиапространстве. С. В. Венидиктов указывает, что «значение различных форм межкультурной коммуникации увеличивается в условиях глобализации социально-политического пространства, унификации общепринятых культурных образцов и практик, появления «глобальных стандартов человечности». Одним из очевидных последствий этого процесса является утверждение тенденции к виртуализации социального взаимодействия в межкультурном пространстве – переноса значительной части традиционных физических практик в сетевое, цифровое, виртуальное пространство» [5, с. 612]. По мнению ряда авторов (Добросклонская, 2021, Самкова, 2021) [6; 7], цифровизация усиливает не только доступность контента, но и вероятность фреймовых искажений и медиапрезентаций при переходе между языками и культурами.

Понятие фрейма восходит к работам Ч. Филлмора, который рассматривал его как структурированную систему знаний, активируемую при восприятии определённых лексем или ситуаций. В рамках фреймовой семантики каждое слово вызывает определённый ментальный сценарий, связанный с нашим опытом и знанием мира [8]. Дж. Лакофф в свою очередь подчёркивал, что фреймы не просто структурируют знания, но и направляют интерпретацию: они задают рамки мышления, метафоры и идеологические установки [9]. Т. Ван Дейк расширил это понятие в контексте дискурсивного анализа, определив фрейм как когнитивную модель, обеспечивающую устойчивость смыслов в коммуникативных ситуациях, особенно в массовой информации. Он подчёркивал, что фреймы участвуют в формировании и воспроизводстве власти, идеологии и социальных норм [2].

В медиадискурсе понятие фрейма приобретает особую значимость, формируя основу так называемых медиафреймов. Определения медиафрейма варьируются в зависимости от подхода исследователей. Классическое определение восходит к дефиниции Р. Энтмана: «To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation,

moral evaluation, and/or treatment recommendation» [10, с. 52]. Согласно когнитивному подходу, фрейм – это «когнитивная схема, которая позволяет классифицировать информацию и интерпретировать события определенным образом». Представители конструктивистского подхода, У. Гамсон и А. Модильяни анализируют основные сюжетные линии, метафоры, крылатые выражения, визуальные образы, нравственные мотивы и другие символические приемы, характеризующие медиадискурс [7, с. 158]. Таким образом, медиафрейм включает выбор темы, постановку проблемы, выделение «виновных» и предполагаемых решений, одновременно задавая эмоциональный и идеологический вектор восприятия. Медиафреймы обеспечивают когнитивную экономию, но в то же время могут способствовать искажению или редукции смыслов, особенно в межкультурной коммуникации, где фреймовые структуры не совпадают.

Проблема усугубляется в условиях цифровой среды, где распространение информации ускорено, а алгоритмы платформ часто усредняют, обрезают или «расплющивают» культурные особенности в пользу универсальности подачи [11]. Это приводит к тому, что медиафреймы, сформированные в одной культуре, теряют свою фреймовую насыщенность при трансляции в другое медиапространство – особенно в автоматических переводах, рерайтах и платформенных адаптациях.

Фреймовые искажения при переходе между языками особенно заметны при попытке интерпретировать национально окрашенные метафоры, встроенные в повседневный медиадискурс. Так, выражение *надеть белое пальто*, широко употребляемое в русскоязычных социальных сетях, апеллирует к фрейму демонстративной моральной непогрешимости и социальной дистанции: «...*белое пальто* является одним из маркеров стремления к конфликтности, служащим для реализации таких интенций пользователя, как ирония, насмешка, а также для указания на скрытую агрессию собеседника» [12, с. 75]. При переводе или восприятии вне русскоязычного контекста метафора либо теряет когнитивную нагрузку, либо искажается вплоть до противоположного значения.

Рассмотрим подробнее историю возникновения и последующие трансформации, произошедшие со ставшим популярным до мемов сочетанием *cancel culture*. Следует отметить, что *cancel culture*, изначально функционирующая как общественно-социальный концепт, в медиадискурсе приобретает черты фрейма. В процессе интерпретации в текстах СМИ она (культура отмены) обрастает устойчивыми ролями (инициатор, объект, аудитория), сценариями (обвинение → публичная реакция → социальные последствия) и ценностными метками (справедливость, травля, протест), как например в статье Д. Ю. Данишимаева, где *cancel culture* рассматривается в качестве новой социальной реальности, проявляющейся через сетевую этику и медиасреду, формирующую горизонтальные нормы общественных ценностей [13]. Таким образом, концепт *cancel culture* актуализируется как медиафрейм,

репрезентирующий моральную и политическую структуру восприятия в конкретной культурной среде. На текущий момент словосочетание закреплено в «Медиясловаре ключевых слов текущего момента» (далее «Медиясловарь») [14], а изначально, в англоязычной медиакультуре он входил в поле фрейма «этическое давление», «ответственность». Появлению термина мы обязаны делу Харви Ванштейна, американскому кинопродюсеру, обвиненному в домогательствах, в результате чего он подвергся публичному осуждению и бойкоту. Он был *canceled* – ‘отменен’, забыт, низвергнут с пьедестала. Этому исходному значению английского глагола *cancel* соответствует первое определение в Медиясловаре: 1. последовательное игнорирование в цифровой среде человека или общности людей в результате общественного осуждения их действий, поступков, слов, что выражается в отказе от публичного освещения их действий, а также в отказе от любой поддержки того, кто подвергнут культуре отмены [15, с. 96]. В статье The Telegraph от 01 января 2025 г. статья под заголовком ‘*Trump’s former lawyer calls for Bank of America investigation amid ‘cancel culture’ row*’ [16] – «Бывший адвокат Трампа призвал провести расследование в отношении Bank of America на фоне скандала вокруг «культуры отмены», где сообщается, что Джон Истман, бывший адвокат президента США Дональда Трампа подвергся гонениям вследствие несовпадения его политических взглядов с последним. В заглавии уже прослеживается первое значение термина *cancel culture* – «культура отмены» из Медиясловаря [14]. Но это только «верхушка айсберга», далее в тексте статьи термин *cancel culture* получает метафорическое расширение и идеологическое окрашивание: ‘*John Eastman said he has still not been given an explanation for why his accounts of 40 years were closed by the banking giant, but believes it was an example of “cancel culture brass in the corporate boardroom”*’. Перевод: «Джон Истман заявил, что ему до сих пор не дали объяснений, почему его счета, обслуживающиеся уже в течение 40 лет, были закрыты банковским гигантом, но он считает, что это был пример культуры отмены, царившей в корпоративном совете директоров». Атрибутивная конструкция *cancel culture brass* может быть переведена как «культура отмены в лице высокопоставленных представителей корпорации», а словосочетание *corporate boardroom* представляет собой метонимию, обозначающую корпоративную элиту, управляющие круги, совет директоров и т.д. Автор статьи переосмысливает *cancel culture* не как инициативу снизу (публики), а как инструмент давления сверху – со стороны власти, корпоративного истеблишмента, который использует «отмену» для устранения неудобных. Таким образом, в англоязычном тексте *cancel culture* приобретает расширенное метафорическое значение. Вместо традиционного понимания «культуры бойкота» снизу (от аудитории, потребителей), используется выражение *cancel culture brass in the corporate boardroom*, где акцент сдвигается на элитарное, системное приме-

нение идеологии отмены, исходящее от управленцев. Такая формулировка усиливает риторический эффект: *cancel culture* предстает не как форма социальной ответственности, а как инструмент давления и исключения в корпоративной политике.

Перейдем ко второму значению термина *cancel culture* в Медиясловаре, его реализации в оригинальном, англоязычном медиадискурсе и возможным интерпретациям: 2. популярное в социальных сетях публичное поведение, нацеленное на отказ человеку или группе в праве на социальную жизнь. Например, в статье онлайн версии газеты The Telegraph под заголовком «*Cancel culture is dead. The mob just hasn't noticed*» от 19 декабря 2024 г. [17] известное выражение обозначает социальную практику, действующую через отказ в праве на социализированное существование, особенно в контексте соцсетей. В названном заголовке наблюдается переход от описания *cancel culture* как мощного общественного инструмента к её иронической девальвации. Употребление фразы *cancel culture is dead* – «культура отмены мертва» в сочетании с образом «толпы, не заметившей смерти» акцентирует риторику пресыщения, утомления от постоянной моральной мобилизации. Здесь проявляется фрейм «утраты остроты», «рутины», а также лёгкая насмешка над самой идеей массового осуждения как «новой нормы». В русскоязычном медиадискурсе фиксируются яркие примеры *cancel culture* в её классическом значении, когда человека или бренд реально исключают из социальных и профессиональных сфер, как например в серии статей, вызвавших широкий общественный резонанс, посвященных Регине Тодоренко и ее «отмене»: «Кто такая Регина Тодоренко и за что её отменяли. Только главное и интересное из биографии. Ведущая популярных телешоу, жена певца и защитница домашнего насилия». «Оправдала домашнее насилие? Все за и против телеведущей Регины Тодоренко» [18, 19].

Нам осталось рассмотреть третье, модифицированное под влиянием текущей политической ситуации значение термина *cancel culture*, представленное в Медиясловаре: 3. последовательные действия западных (медиа)деятелей, направленные на отказ в праве на социализированное существование русскому народу. Третий уровень *cancel culture* встречается в риторике политиков и СМИ, где термин трясут не ради бойкота конкретного человека, а как инструмент геополитического давления: «*President complains west 'trying to cancel a whole 1,000-year culture, our people'... Russian writers and books are now cancelled*» [20]. Перевод: «Владимир Путин прямо заявил, что Запад пытается отменить тысячелетнюю культуру, наш народ». Здесь фрейм расширен до уровня нации, отменяют все, что связано с русской культурой и людьми. Пример того, как западные СМИ говорят об отмене России: «*We are talking about attempts to destroy our state – its 'cancellation'... Russia is now being canceled*» – «Мы говорим о попытках разрушить наше государство – о его отмене», где фрейм «мораль, нравственные принципы» акцентируется не на отдельном человеке, а на всем государстве [21].

Что касается СМИ Узбекистана, в проанализированных источниках, таких как информационно-аналитический сайт Asia Terra, радио Ozodlik.uz, онлайн издание Газета.uz получило широкий общественный резонанс дело узбекского видеоблогера Отабека Саттори, в 2021 г. осуждённого на 6,5 лет тюрьмы по обвинениям в вымогательстве и клевете – фактически за критику местных чиновников [22]. Здесь фрейм «социального исключения», когда человек физически и медийно «отключён» от общения и общественной жизни, без формального института *cancel culture*, по сути транслирует тот же эффект «отмены». В дальнейшем обвинения стали моделью для уголовного преследования других блогеров [23].

Таким образом, можно отметить, что в узбекоязычном медиапространстве фрейм *cancel culture* пока не получил явного распространения. Это может быть объяснено сочетанием факторов: традиционно-нормативным характером подачи информации, ограниченностью публичной критики, отсутствием института репутационного «краха», а также доминированием этико-ориентированных моделей коммуникации. В отличие от англоязычного и частично русскоязычного медиадискурса, где фрейм *cancel culture* оформляется как самостоятельное явление, в узбекских СМИ подобные процессы реализуются в иных формах – скорее как социальное порицание, мягкое дистанцирование, либо нормативно-религиозная реакция, не оформленная в виде устойчивого медиафрейма.

Проведённый фреймовый анализ медиатекстов трёх языковых и культурных пространств – англоязычного, русскоязычного и узбекоязычного – позволил выявить особенности репрезентации и интерпретации одного из значимых современных понятий цифровой эпохи – *cancel culture*.

В англоязычном медиадискурсе *cancel culture* демонстрирует широкую семантическую эволюцию: от общественной моралистской реакции (бойкот публичных фигур) до ироничной критики самой идеи отмены как уставшей, ритуализированной практики. Термин активно фреймуется как инструмент социального давления снизу, а также, в более поздних метафорических употреблениях, – как проявление корпоративной или идеологической цензуры сверху.

В русскоязычных источниках наблюдается как перенос англоязычного фрейма, так и его идеологическое перераспределение. *Cancel culture* здесь всё чаще трактуется как угроза российской культуре, формируя третий уровень фрейма – «отмену народа». Это отражает особенности политического и культурного контекста, где концепт приобретает черты идеологической борьбы.

В узбекоязычном медиапространстве концепт *cancel culture* как устойчивый фрейм отсутствует, что может рассматриваться как фреймовая лакуна, обусловленная традиционностью дискурса, низкой публичностью моральных санкций и доминированием официально-нормативного стиля коммуника-

ции. Тем не менее отдельные кейсы преследования блогеров, критикующих власть, можно рассматривать как функциональный аналог «отмены», реализующийся в иной, институционально-закрытой форме.

Таким образом, межкультурный анализ медиафреймов позволяет зафиксировать как семантические трансформации одного и того же концепта в разных медиакультурах, так и разрывы в когнитивных моделях, обусловленные идеологией, медийной структурой и культурной нормой интерпретации. Это подчёркивает актуальность фреймового подхода в исследовании цифровой межкультурной коммуникации.

ЛИТЕРАТУРА

1. Дугин Е. Я. Медиаиндустрия. М. : Канон РООИ «Реабилитация», 2021. 416 с.
2. Т. А. ван Дейк Дискурс и власть: Репрезентация доминирования в языке и коммуникации. М. : Книжный дом «Либроком», 2013. 344 с.
3. Шорова М. Д., Бориева М. К., Текуева А. А. Роль средств массовой информации в развитии межкультурной коммуникации в условиях цифровизации [Электронный ресурс] // Филология : научные исследования. 2024. № 2. С. 38–49. DOI: 10.7256/2454_0749.2024.2.69924 EDN: VRSZCP URL: https://nbpublish.com/library_read_article.php?id=69924 (дата обращения 05.07.2025).
4. Денисова Г. В., Смирнова О. В., Сапунова О. В. Лингвокультурные универсалии как доминирующий фактор восприятия мемов [Электронный ресурс] // Филологические науки. Язык. Культура. Социум. 2023. № 6. С. 43–48. DOI 10.20339/PhS.6-23.04. (дата обращения 10.07.2025).
5. Венидиктов С. В. Межкультурная коммуникация в условиях виртуализации социального взаимодействия [Электронный ресурс] // CyberLeninka.ru. Режим доступа: <https://cyberleninka.ru/article/n/mezhkulturnaya-kommunikatsiya-v-usloviyah-virtualizatsii-sotsialnogo-vzaimodeystviya> (дата обращения: 06.07.2025).
6. Добросклонская Т. Г. Лингвомедийное конструирование события в новостном дискурсе // Медиалингвистика. Вып. 8. Язык в координатах массмедиа : материалы V Междунар. научн. конференции (Санкт-Петербург, 30 июня – 2 июля 2021 г.) / науч. ред. Л. Р. Дускаева, отв. ред. А. А. Малышев. – СПб. : Медиапапир, 2021. С. 108-112.
7. Самкова М. А. Фреймирование в опыте конструирования образа города // [Электронный ресурс]. Political Linguistics. 2021. No 5 (89). P. 157-167. DOI 10.26170/1999-2629_2021_05_18 (дата обращения 03.07.2025).
8. Fillmore Ch. J. Frame Semantics. In Linguistics in the Morning Calm: Selected Papers from SICOL-1981. Seoul: Hanshin Publishing Co. 1982. P. 113-137.
9. Lakoff G. Don't Think of an Elephant! Know Your Values and Frame the Debate. White River Junction, VT: Chelsea Green Pub. Co., 2004. 124 p.
10. Entman R. M. Framing: Toward clarification of a fractured paradigm. Journal of Communication. 1993. 43(4), P. 51–58.
11. Manovich L. Software Takes Command: Extending the Language of New Media. New York; London: Bloomsbury Academic. 2013. 376 с.

12. Абыякая О. В. «Белое пальто» в интернет-коммуникации: особенности функционирования // Медиалингвистика. Вып. 11. Язык в координатах массмедиа: материалы VIII Междунар. науч. конф. (Санкт-Петербург, 26–29 июня 2024 г.) / науч. ред. Л. Р. Дускаева, отв. ред. А. А. Малышев. – СПб. : Медиапир, 2024. С. 74–77.
13. Дашиинимаев Г. Ю. «Cancel culture как новая социальная реальность» // Вестник Бурятского государственного университета. Философия. 2021. № 1. С. 19–24. – DOI: 10.18101/1994-0866-2021-1-19-24 (дата обращения: 13.07.2025).
14. Культура отмены // DataSlov: Словарь новых слов русского языка. – URL: <https://dataslov.ru/word/kultura-otmeny/> (дата обращения: 11.07.2025).
15. Прокофьева Н. А., Щеглова Е. А. «Медиасловарь ключевых слов текущего момента» в реальности цифровой среды // Медиалингвистика. Вып. 11. Язык в координатах массмедиа: материалы VIII Междунар. науч. конф. (Санкт-Петербург, 26–29 июня 2024 г.) / науч. ред. Л. Р. Дускаева, отв. ред. А. А. Малышев. – СПб. : Медиапир, 2024. С. 94–97.

ИСТОЧНИКИ

16. Trump’s former lawyer calls for Bank of America investigation amid ‘cancel culture’ row // The Telegraph. – 01.01.2025. – URL: <https://www.telegraph.co.uk/us/news/2025/01/24/john-eastman-urges-trump-investigate-bank-of-america/> (дата обращения: 13.07.2025).
17. Cancel culture is dead. The mob just hasn’t noticed // The Telegraph. – 19.12.2024. – URL: <https://www.telegraph.co.uk/films/0/cancel-culture-is-dead-the-left-hasnt-noticed/> (дата обращения: 13.07.2025).
18. Рудич К. Кто такая Регина Тодоренко и за что её отменяли. Только главное и интересное из биографии // Secretmag. 28.05.2023 - URL: <https://secretmag.ru/enciklopediya/kto-takaya-regina-todorenko-i-za-chto-eyo-otmenyali-tolko-glavnoe-i-interesnoe-iz-biografii.htm> (дата обращения: 12.07.2025).
19. Оправдала домашнее насилие? Все за и против телеведущей Регины Тодоренко // Москва 24. – 27.04.2020. – URL: <https://www.m24.ru/articles/obshchestvo/27042020/156651> (дата обращения: 12.07.2025).
20. Putin complains west ‘trying to cancel a whole 1,000-year culture, our people’ // Euronews (via Reuters). – 26.03.2022. – URL: <https://www.euronews.com/2022/03/26/us-ukraine-crisis-russia-culture> (дата обращения: 13.07.2025).
21. Lewis H. Of Course Putin Is Being Canceled // The Atlantic. – 08.03.2022. – URL: <https://www.theatlantic.com/ideas/archive/2022/03/russia-ukraine-war-cancel-putin-culture-war/626978/> (дата обращения: 13.07.2025).
22. Чем закончилась история сурхандарьинского блогера Отабека Саттори. – 21.07.2024. – URL: <https://rus.ozodlik.org/a/32806097.html> (дата обращения: 12.07.2025).
23. Узбекистан: Блогер приговорен к 8 годам тюрьмы. – 06.12.2023. – URL: https://ihahr.org/news/uzbekistan-blogger-prigovoren-k-8-godam-tyurmy?utm_source=chatgpt.com (дата обращения: 12.07.2025).

УДК 81'42

Силанов Еремей Александрович
МГУ им. М. В. Ломоносова
г. Москва, Россия

Eremey Silanov
Lomonosov Moscow State University
Moscow, Russia
silanoveremey@gmail.com

МЕДИАСТИЛИСТИЧЕСКИЙ АНАЛИЗ ЦИФРОВЫХ ИГР КАК РЕЧЕВОГО ПРОСТРАНСТВА

(на примере российских игр «Смута», «Atomic Heart» и «Black Book»)

Статья посвящена медиастилистическому анализу цифровых игр как полимодальных медиатекстов, формирующих особое речевое пространство. Исследование основано на теоретических подходах Т. Г. Добросклонской, Н. И. Клушиной и Б. Тошовича, представляющих различные модели медиастиля и медиатекста. Особое внимание уделено понятию стилевой доминанты как ключевому элементу организации игрового дискурса. Анализируются речевые регистры, идиостили персонажей, экспрессивные и лексико-синтаксические особенности в играх «Смута», «Atomic Heart» и «Black Book». Показано, что текст в цифровой игре представляет собой креолизованную структуру, в которой вербальные, визуальные и интерактивные элементы взаимосвязаны и ситуативно обусловлены. Делается вывод о продуктивности медиастилистического подхода для анализа игровых текстов, отражающих исторические и коммуникативные особенности.

Ключевые слова: медиастилистика; цифровая игра; игровая речь; стилистические доминанты; интертекстуальность; креолизованный текст.

MEDIASTYLISTIC ANALYSIS OF DIGITAL GAMES AS A SPEECH SPACE (based on the Russian games “Smuta”, “Atomic Heart” and “Black Book”)

The article is devoted to the media stylistic analysis of digital games as polymodal media texts that form a special speech space. The study is based on the theoretical approaches of T. G. Dobrosklonskaya, N. I. Klushina and B. Toshovich, who present different models of media style and media text. Special attention is paid to the notion of style dominance as a key element in the organization of game discourse. Speech registers, idiostyles of characters, expressive and lexico-syntactic features in the games “Smuta”, “Atomic Heart” and “Black Book” are analyzed. It is shown that the text in a digital game is a creolized structure in which verbal, visual and interactive elements are interrelated and situationally conditioned. It is concluded that the media stylistic approach is productive for analyzing game texts reflecting historical and communicative features.

Key words: media stylistics; digital game; game speech; stylistic dominants; intertextuality; creolized text.

Современные цифровые игры формируют новое пространство речевой активности, в котором язык утрачивает строгое разграничение на устное и письменное, а текст становится интерактивным, многослойным и ситуативным. Анализ такого текста требует теоретических оснований, способных учитывать лексико-грамматические особенности контента, его полимодальность, визуальную структуру, ролевую обусловленность и коммуникативную направленность. Этим задачам отвечает медиастилистика – дисциплина, возникающая на стыке стилистики, медиалингвистики и дискурсивного анализа.

Понятия «медiateкст» и «медиастиль» стали центральными в рамках этой научной парадигмы. По определению Т. Г. Добросклонской, медiateкст – это речевое произведение, предназначенное для трансляции в медиасреде, обладающее сложной семиотической природой, включающее визуальные, вербальные, аудиальные и интерактивные компоненты [1, с. 28]. Такой текст структурируется через взаимодействие кодов и требует от адресата особой компетенции – способности считывать смыслы на нескольких уровнях.

Н. И. Клушина определяет медиастиль как динамическую речевую форму, возникающую на основе интенции, жанровых ожиданий и стилистической установки говорящего [2, с. 69–77]. В её подходе медиастиль – это инструмент воздействия, структурирующий восприятие текста и формирующий эмоционально-смысловую рамку. Особенность медиастиля в его ситуативной гибкости и зависимости от среды: стиль здесь вырабатывается в процессе взаимодействия участников коммуникации.

Одним из поворотных моментов в развитии медиастилистики стало формирование парадигмы интернет-стилистики, предложенной Б. Тошовичем. Он рассматривает медiateкст как структуру гипертекста – децентрализованную, нелинейную, креолизованную, в которой исчезает понятие авторского центра, а пользователь становится «числителем» – одновременно читателем и автором [3, с. 64–67]. Такая модель требует новых описательных категорий: текст становится сетевой конфигурацией смыслов, активизируемой в ходе взаимодействия.

Сравнение трёх моделей медиастилистики (Клушиной, Добросклонской и Тошовича) позволяет зафиксировать смещение научного интереса от анализа формы к исследованию интенции и среды речевой деятельности. Клушина сосредотачивается на речевых стратегиях и идиостиле, Добросклонская – на системной организации медиадискурса, Тошович – на деструктурированном цифровом тексте.

Для анализа цифровых игр медиастилистика оказывается особенно продуктивной, поскольку игровая речь неотделима от игровой графики, геймплея. Игровой текст реагирует на действия игрока, он «живет» в интерактивной ситуации и требует понимания как сказанного, так и визуально оформленного. Это делает стилистический анализ таких текстов невозможным без медиастилистической оптики. В центре внимания оказываются идиостили персонажей, способы переключения регистров, экспрессивные приёмы и речевые маски, формирующие мир игры как речевую и культурную модель.

Цифровая игра функционирует как полимодальный медiateкст, в котором вербальное содержание встроено в аудиовизуальную и интерактивную структуру. Текст здесь создаётся на стыке замысла разработчика и действий игрока. Это совокупность текстов, объединённых общей тематикой и функционирующих внутри компьютерной игры различных жанров [4].

Семантика игрового текста складывается слов, графики, звуков, мимики персонажа. Эти элементы работают совместно, формируя креолизованное

пространство высказывания [5]. Язык в игре выполняет информативную, навигационную функцию: он помогает ориентироваться, реагировать, принимать решения [6].

Речевой портрет персонажа встроен в игровой мир. Его стиль зависит от жанра, социального положения, эмоционального состояния, что делает речь мощным инструментом типизации [7]. Каждая реплика в игровом пространстве – это одновременно действие. Она запускает сцены, изменяет ход событий, создаёт эффект выбора [8]. Таким образом, цифровая игра выстраивает речь как активный модуль повествования.

Анализ цифровой игры как речевого пространства невозможен без понятия стилевой доминанты – устойчивого стилистического признака, формирующего общую тональность медиатекста. В рамках медиастилистики стилевые доминанты определяются как узлы смыслового и экспрессивного напряжения, на которых держится структура игрового текста. По определению Н. И. Клушиной и А. В. Николаевой, они одновременно объединяют тексты, циркулирующие в одном медиапространстве, и позволяют дифференцировать речевые коды разных жанров и форм [9, с. 73]. В отличие от дискурсивных доминант, сосредоточенных на темах и когнитивных установках, стилевые фокусируются на лексике, синтаксисе, интонации и речевой стратегии.

Над определением особенностей стилевых доминант много и плодотворно работал С. Т. Золян, связывая их с идиостилем как системой приёмов, организующих текст на глубинном уровне. Доминанта – это способ изменить обычные функциональные отношения между единицами текста, наделять их художественным, экспрессивным, авторским акцентом [10]. В игровом дискурсе эта категория становится особенно важной, поскольку текст здесь описывает и запускает действия, реагирует на выбор игрока.

Стилевые доминанты в цифровых играх можно условно классифицировать по четырём направлениям. Во-первых, это лексические маркеры эпохи. В исторических играх, например, таких как «Смута», широко используются архаизмы, историзмы, религиозная лексика, формулы официального дискурса. Речь боярина, священника или крестьянина оформляется по-разному, с учётом ролевой функции и культурного контекста.

Вторую группу составляют экспрессивные доминанты. Это эмоционально насыщенные конструкции, парцелляции, резкие переходы интонации, речевые сбои. Такие элементы маркируют пики напряжения, придают реплике ритм, создают эффект «живой речи». Например, использование междометий, разорванных синтаксических структур или повторов.

Отдельно стоит выделить разговорные формы и сниженные регистры, включая просторечие, диалектизмы, неологизмы, окказионализмы. Персонажи с нарочито грубой или сленговой речью вызывают у игрока эффект узнавания и сопричастности. Просторечие перестаёт быть маркером социального низа и становится речевой стратегией эмоционального вовлечения. Оно сигнализирует, что персонаж «свой», что его речь – это реакция на игровую ситуацию.

В одном и том же диалоге могут одновременно звучать разные речевые регистры. Персонаж способен начать реплику, используя официальные этикетные формы и тут же, в границах одного предложения, использовать грубо-просторечную лексику. Такая гибкость создает многослойный текст, в котором сталкиваются разные стили, маски, роли.

Важно подчеркнуть, что речь в игре имеет свою функцию. Реплика выполняет действие: убеждает, провоцирует, оценивает. В этом смысле стилиевые доминанты – это и средства смыслообразования. Их выбор определяется интенцией персонажа и логикой сцены. Эмоционально окрашенные лексемы усиливают драматизм или создают юмористический эффект, в зависимости от контекста и целей.

Таким образом, стилистические доминанты создают узнаваемый стиль игры, управляют восприятием, структурируют взаимодействие с игроком.

В цифровой игре «Смута» речевая ткань представляет собой сложную систему стилистических доминант, направленных на воссоздание исторической достоверности. Стилиевой каркас формируется через архаическую, религиозную и просторечную лексику, а также за счёт выраженной оценочности.

Центральное место занимает лексико-семантический пласт историзмов и архаизмов: слова «челядь», «изволь», «рекрут», «дьяк», «вельможа», «писец», «казна» или «ратник» выступают как действенные речевые механизмы, участвующие в построении сюжетных поворотов. Их функция реконструктивная – задающая атмосферу XVII в.

Грамматические архаизмы (инверсии, специфический порядок слов, архаичная морфология) – «аще», «сия», «чаял», «внемли», «доколе» – дополняют общее звучание игры. В речевом регистре монахов, странников, старцев слышна интонация проповеди: «Даже в самой чёрной душе луч света найтись может», «Господи, вразуми», «всемилоостивый», «божья воля». Так формируется высокостилевое звучание, стилизованное под церковнославянскую традицию.

При этом рядом активно функционирует просторечная и сниженная лексика: «харя», «жратва», «мироед», «пузо», «брешет», «дочирикаетесь». Эти элементы активны в речи казаков, солдат, крестьян. Они формируют так называемую речевую ландшафтность, где каждая социальная или территориальная группа говорит на «своём» языке: «Пуцай князь решает, нам бы лишь сдюжить».

Игровая речь строится также через оценочные конструкции, насыщенные метафорами, фразеологизмами, фольклорными образами: «Ты, пес смердячий!», «Псы злые, но голодные – кинешь колбаски, и лаять перестанут», «Ты на пузо мое не смотри – с голодухи надуло». Эти фразы стилистически маркированы, задают эмоциональное поле, программируя поведение других персонажей и самого игрока.

Наконец, важной стилевой доминантой является полифония, выстроенная на иерархическом различии речевых стратегий: бояре говорят, используя книжную лексику, монахи – архаично, солдаты – резко и грубо. Такая системная «разноязычность» формирует многослойный медиатекст, в котором каждое слово и каждая синтаксическая форма выполняют повествовательную и идентификационную функцию. Здесь язык – это и историческая маска, и инструмент действия.

Цифровая игра «Atomic Heart» разворачивает своё повествование в ретрофутуристической антиутопии, построенной на советской эстетике. Через язык в игре прослеживается конфликт между «вежливой» утопией и жестокой реальностью. На этом противоречии строятся её стилистические доминанты.

Первым и наиболее заметным приёмом становится контраст речевых регистров. Главный герой, майор П-3, использует сниженную, разговорную, а порой обценную лексику, резко противопоставленной официальному языку роботов и учёных. В то время как автоматизированная система может говорить технократическим языком, используя слова вроде: «Инъекция нейрополимера, мнемосфера», П-3 отвечает: «Мнемо–что? Ты вообще по-русски умеешь, кусок жести?».

Следом выделяется технократический и идеологический канцелярит, в котором живут роботы, научные сотрудники. Такие лексемы, как «нейрополимер», «Контур Коллектив», «мнемосфера», «операторная комната перераспределения ресурсов», «гравитационная колба», – создают ощущение логичности и закрытости мира, в котором язык структурирует реальность, но не объясняет её.

Показателен и лексический пласт советизмов, «инженерный гений Сеченова», «во благо советского народа», «дом Советов», «передовой работник» – всё это звучит с механической точностью, особенно в устах робота Терешковой: «Ваше благополучие – приоритет предприятия».

Доминантой становится метафоризация технологического и политического языка: «... вонзи клинок знаний в висок капиталистической бюрократии».

В этой игре также в полной мере проявляется полифоничность игрового текста: каждый персонаж говорит на своём языке – роботы и системы демонстрируют идеальную бюрократическую речь, герой П-3 разрушает её снижением и агрессией.

Цифровая игра «Black Book» формирует речевое пространство, насыщенное фольклорно-религиозными кодами. Здесь язык – механизм эмоционального погружения. В стилистической архитектуре игры доминируют архаизмы, просторечия, сказовые синтаксические конструкции, а также лексика, связанная с сакральной и демонической онтологией.

Первым слоем является сакрально-религиозная лексика, органично встроенная в повседневные диалоги. Формулы вроде «Грех великий на душу берёшь», «Не бросать же душу христианскую», «Думаю, надо молебн

заказать» становятся маркерами духовной координатной сетки, в которой вера – бытовая норма поведения. Эта норма прослеживается и через именование разных предметов вроде: «полотенце церковное», «псалтырь», «икона адописная».

Наряду с этим активно функционируют просторечия: «Ты че, девка, в уме ли?», «Бисячина мохнатая!», «Ох ты ж, гниды у тебя заваренные!». Эти фразы создают плотный интонационно-эмоциональный фон.

Особое место занимает сказовая интонация – стилистическая доминанта, формирующая речь. Синтаксис построен на сбоях, повторах, инверсиях: «Погоди-погоди!», «Сгинь с глаз моих!», «Заговор читать начну, а черт блазнить будет». Эти конструкции имитируют устную речь рассказчика.

Среди лексических доминант выделяются зооморфизмы и инвективы: «дрянь ты московская», «прощелыга», «жмурики», «черт его неси». Персонажи через язык создают границы между «своим» и «чужим».

В «Black Book» стилевые доминанты игры демонстрируют живую устную традицию.

Необходимо особо отметить, что язык в играх, особенности построения персонажных диалогов не менее важны, чем графическая их составляющая.

Образовательная и просветительская функция цифровых игр несомненна. Цифровые игры становятся важным элементом образовательной экосистемы XXI в. Их сила – в способности превращать абстрактные знания в личный опыт, а сложные темы – в увлекательное интерактивное путешествие. Например, в прошлое, что прослеживается в игре «Смута». Популярность подобного контента подтверждается статистикой: игру «Смута» и учебные приложения «Смутное время» от студии «Сайберия Нова» скачали более миллиона раз [11]. Это достижение дало разработчикам возможность получить государственное финансирование на продолжение работы. Игра помогает вспомнить исторические имена и реалии, получить достоверные сведения об эпохе, чему способствуют виртуальные экскурсии, справочные материалы, объяснения архаизмов и историзмов, внимание к визуальной среде (архитектура, костюмы, предметы быта).

Важно развивать отечественные цифровые игры для детей и подростков, в которых будет задействован обучающий потенциал медиаречи. Такие игры могут стать коммуникативными тренажёрами, инструментами социальной адаптации, средствами воспитания. Исследования показывают, что игры развивают многозадачность, эмоциональный интеллект, эмпатию, пополняют словарный запас. Работа лингвистов и стилистов над игровыми сценариями пока не институционализирована, хотя необходимость в этом уже очевидна.

ЛИТЕРАТУРА

1. Добросклонская Т. Г. Медиалингвистика: системный подход к изучению языка СМИ. Москва : URSS, 2008.
2. Клушина Н. И. Медиастилистика // Экология языка и коммуникативная практика. 2014. № 2. С. 69–77.
3. Тошович Б. Интернет-стилистика. Москва : ФЛИНТА: Наука, 2015. С. 64–67.

4. Третьякова М. Ф., Титов В. В. Особенности реализации языковой игры в компьютерном игровом дискурсе // Амурский научный вестник. 2023. № 4. С. 108–112.
5. Саяхова Д. К. Семиотическая природа компьютерного дискурса (на примере видеоигр) [Электронный ресурс] // Вестник Башкирского государственного университета. 2021. № 1. С. 250–256. DOI: 10.33184/bulletin-bsu-2021.1.39.
6. Никонова Е. В. Функции языковой составляющей в компьютерной игре // Вестник МГЛУ. Гуманитарные науки. 2019. № 6 (822). С. 146–151.
7. Перфилов Ю. А. Языковые особенности текстов компьютерных игр // Гуманитарные науки. 2019. № 5. С. 188–190.
8. Овечкина Е. А., Васильева С. С. Речевая характеристика персонажа как средство создания художественного образа // Вестник Волгоградского государственного университета. Сер. 8: Литературоведение. 2019. № 1 (18). С. 20–21.
9. Клушина Н. И., Николаева А. В. Введение в интернет-стилистику: учебник. Москва : ФЛИНТА, 2020. С. 73.
10. Золян С. Т. От описания идиолекта – к грамматике идиостиля // Язык русской поэзии XX века. Москва : Ин-т русского языка АН СССР, 1989. С. 238–259.
11. Игру "Смута" и учебные приложения на ее основе скачали 1 млн раз // Тасс URL: <https://tass.ru/ekonomika/22507137> (дата обращения: 12.05.2025).

УДК 81'373.45

Хрущева Оксана Александровна

кандидат филологических наук,
доцент кафедры английского
и китайского языков

Оренбургский государственный
университет
г. Оренбург, Россия

Oxana Hrushcheva

PhD in Philology, Associate Professor
at English and Chinese Languages

Department

Orenburg State University

Orenburg, Russia

hrox@mail.ru

СИНКРЕТИЗМ ПОЛИМОДАЛЬНЫХ ТЕКСТОВ В СОВРЕМЕННОЙ ИНТЕРНЕТ-КОММУНИКАЦИИ

В статье рассматриваются актуальные особенности интернет-коммуникации как наиболее востребованного формата общения для большинства принимающих участие в исследовании респондентов, а именно превалирование развлекательного контента над познавательным, доминирование визуального содержания над текстовым, предпочтение видео-формата в виде коротких фрагментов. В качестве материала для исследования избран ряд мемов, получивших широкое распространение среди интернет-пользователей в 2025 г. Методы, применяемые в ходе работы, включают контент-анализ, интерпретацию, описание, классификацию. Автор акцентирует внимание на мемах, сгенерированных искусственным интеллектом и отнесенных интернет-сообществом к категории «брейнрот» (о чем свидетельствует соответствующая разметка подобных постов). Данное наименование и хэштег используются для обозначения информации низкого интеллектуального уровня и примитивного качества. На наглядных примерах демонстрируется синкретический характер интернет-текстов, в которых наличествуют визуальный и текстовый компоненты. Автор ставит под сомнение наличие семантической и семиотической ценности у подобных мемов несмотря на их массовое распространение.

Ключевые слова: интернет-мем; синкретизм; полимодальный текст; поликодовость; искусственный интеллект.

SYNCRETISM OF POLYMODAL TEXTS IN PRESENT-DAY INTERNET COMMUNICATION

The article unveils present-day features of internet communication as the most popular means to interact for the vast majority of people having taken part in the research. These features include more entertaining content than educational posts, more visual effects than printed texts, more video clips of minimum form. The paper deals with the topical Internet memes that have become viral among users since early 2025. The methods employed in the study include content analysis, interpretation, description and classification. The author focuses upon the analysis of the AI-generated memes that are categorized as brain rot (which is a hashtag accompanying the relevant posts). This term is used to indicate the information of poor quality and low brow. The article studies some common memes that can be characterized as syncretic due to the presence of visual and textual components in their structure. The author expresses some doubt that this kind of memes has semantic and semiotic value notwithstanding their mass circulation.

Key words: an internet meme; syncretism; polymodal text; polycode; artificial intelligence.

В современном мире интернет-коммуникация является доминирующей для подавляющего большинства представителей различных социальных

групп, что обуславливает ее динамическое развитие и всеобъемлющее пространство. Социологи, психологи и лингвисты на протяжении десятилетий проводят всеобъемлющие исследования данного типа общения и взаимодействия и приходят к выводу о том, что интернет-коммуникация обладает набором особенностей.

И. А. Якоба в своих научных работах формулирует список отрицательных, положительных и нейтральных характеристик исследуемого феномена. По мнению исследователя, «к положительным коммуникативным характеристикам относятся следующие:

- передача большего количества информации за минимальное количество времени;
- возможность передачи информации при недоступности адресата;
- необходимость лаконично и четко формулировать свои мысли;
- одновременная передача информации и эмоций (при помощи определенных символов);
- число собеседников не ограничено, они могут вступать в беседу и покидать её на любом этапе;
- возможность подключить изображение и звук, представлять фотографии и видеосюжеты, общаться в режиме реального времени и т.п. к тексту на естественном языке;
- добровольность и желаемость контактов;
- возможность совместить коммуникацию и автокоммуникацию: тексты, посылаемые другому, одновременно становятся доступны как адресату, так и отправителю.

К отрицательным коммуникативным характеристикам отнесены:

- затруднение в понимании содержания при большом количестве неизвестных сокращений и других специфических символов, присущих электронному письму;
- изобретение особых для интернет-коммуникации специфических способов эмоционального наполнения текста как компенсаторной практики из-за отсутствия визуального контакта;
- пренебрежение правилами грамматики, синтаксиса и орфографии;
- появление ошибок в профессиональной сфере деятельности у людей, привыкших к «лояльному стилю» бытового виртуального общения, например, при ответе на корреспонденцию и т.п.;
- анонимность участников коммуникации (отсутствие внешних характеристик: пол, возраст, внешняя привлекательность или непривлекательность, социальный статус и т.д., отсутствие признаков невербального общения).

Нейтральными коммуникативными характеристиками можно считать:

- дистантность в пространстве и во времени;
- опосредованность;
- наличие гипертекста;

- креолизованность компьютерных текстов;
- по преимуществу статусное равноправие участников;
- специфическая компьютерная этика;
- комбинация различных типов дискурса;
- слабая социальная регламентированность вербального поведения;
- снятие жестких социальных конвенций и культурных границ» [1, с. 367].

Приведенный список характеристик убеждает нас в том, что интернет-коммуникация является многоаспектным, динамично развивающимся явлением современной жизни, вызывающим у участников данного процесса неоднозначные реакции и эмоции. Однако, в данном контексте развитие не должно ассоциироваться напрямую с совершенствованием форм, тактик и стратегий коммуникаций, так как в текущий момент в онлайн-среде наблюдается примитивизация контента, получившая название «брейнрот».

Указанное наименование в 2024 г. приобрело статус «Слово года» в английском языке в связи с повсеместным распространением в сети интернет информации крайне низкого качества и ее систематическим потреблением широким кругом пользователей, что вызывает обеспокоенность в связи с их психическим здоровьем и уровнем интеллектуального развития. В переводе с английского языка *brain rot* означает «гниение мозга», что объективно отражает состояние типичного представителя современного общества, отдающего абсолютное предпочтение интернет-коммуникации как для профессиональных целей, так и в режиме отдыха.

В российском онлайн-сообществе «брейнрот» используется в качестве хештега для разметки интернет-мемов определенного формата: 84 % примеров из корпуса исследования являются полимодальными (поликодовыми) текстами, сочетающими в своей структуре как вербальные, так и невербальные компоненты. Согласно определению Е.С. Шмелевой, «интернет-мем – это полимодальная (или, используя другую терминологию, креолизованная) единица представления информации, поскольку в нем графическая составляющая (а именно, картинка или фото) сопровождается текстовой» [2, с. 175].

Р. Якобсон различал «гомогенные сообщения, которые имеют в качестве основы одну семиотическую систему, и синкретические, которые основываются на комбинации и синтезе различных знаковых систем, например, сообщения с сочетанием аудиальных и визуальных семиотических средств» [3, с. 327–328]. Корпус примеров исследования, посвященного комплексному анализу интернет-мемов, показывает, что синкретичный характер присущ подавляющему большинству отобранных текстов, так как в их структуре синтезируются текстовое содержание, фото-изображение и/или флеш-анимация.

Ранее нашему рассмотрению подвергалась популярная в среде молодежи единица *вайб*, которая в прошлом году была признана словом года в русском языке, а также полимодальные тексты с данным компонентом

в составе [4, с. 508]. Признаем, что такие тексты как «+вайб» (*плюс вайб*), «-вайб» (*минус вайб*) объективно являются содержательными, связными и неразрывно сопряжены с оценкой события или ситуации как со стороны говорящего, так и слушающего, в то время как мемы, появившиеся в интернет-пространстве в текущем году вызывают когнитивный диссонанс и неприятие. Рассмотрим наиболее типичные примеры подобных полимодальных текстов (Рисунки 1–3).



Рис. 1–3. Популярные интернет-мемы (2025 г.)

Интернет-мемы, представленные на рисунках, имеют в своем составе изображения и надписи на латинице; в русскоязычном интернет-сообществе данные иноязычные надписи воспроизводятся с характерной итальянской интонацией, что связано с происхождением приведенных полимодальных текстов. С точки зрения структуры, приведенные примеры имеют отношение к модели концептуальной интеграции, что наблюдается как в вербальных, так и в визуальных компонентах. Однако, с точки зрения смыслового наполнения, отмечается полное отсутствие контекстуальных связей и прагматической составляющей, что отмечают среднестатистические участники интернет-коммуникации, отмечая подобные образцы хештегом #брейнрот.

На фоне предпринятого опыта анализа актуальных интернет-мемов важно обратиться к мнению Е.С. Кубряковой относительно текста и критериев его определения: «текст существует как источник излучения, как источник возбуждения в нашем сознании многочисленных ассоциаций и когнитивных структур (от простых фреймов до гораздо более сложных ментальных пространств и возможных миров). Текст в силу этого свойства показателен именно в том, что из него можно вывести, заключить, извлечь. Он являет собой поэтому образец такой сложной языковой формы, такого семиотического образования, которое побуждает нас к творческому процессу его понимания, его восприятия, его интерпретации, его додумывания – к такого рода когнитивной деятельности, которая имеет дело с осмыслением человеческого опыта, запечатленного в описаниях мира и служащего сотворению новых ступеней познания этого мира» [5, с. 80-81]. Опираясь на данные критерии, мы приходим к выводу, что при констатации полимо-

дального характера современных интернет-мемов и их массового распространения не представляется возможным судить об их семантической и семиотической ценности, ввиду их фактического отсутствия. Более того, сомнению также стоит подвергнуть коммуникативный потенциал приведенных образцов, так как обмен подобными мемами не имеет отношения к рациональному сознанию. Участники интернет-коммуникации используют их в качестве готовых шаблонов без привязки к конкретной ситуации или контексту, выражая непонимание и неопределенные эмоции, следуя языковой моде и интернет-трендам, что свидетельствует о массовом бессознательном отношении к собственной речи и ее содержанию.

Внимание к интернет-коммуникации с актуальными явлениями, присутствующими данной сфере жизнедеятельности современного человека, выступает характерной чертой научных исследований и без сомнения требует применения междисциплинарного подхода в связи со сложным характером наблюдаемых объектов, а также на фоне использования искусственного интеллекта, претендующего стать полноценной заменой природного разума. Явление «брейнрот», несмотря на свое иноязычное происхождение и название, приобретает глобальный характер и охватывает все больше интернет-пользователей, для которых онлайн-среда предлагает замену реальным сущностям в быту, профессии, взаимоотношениях без учета качества, ценности и потенциала. Синкретические компоненты интернет-коммуникации благодаря многоаспектной аудио-видео-визуально-текстовой структуре отвечают принципу карнавализации, что делает общение легким, непринужденным и во многом бездумным.

ЛИТЕРАТУРА

1. Якоба И. А. Особенности интернет-коммуникации (социологический, лингвистический, гендерный аспекты) // Вестник ИрГТУ. 2012. №3 (62). URL: <https://cyberleninka.ru/article/n/osobennosti-internet-kommunikatsii-sotsiologicheskii-lingvisticheskii-gendernyy-aspekty> (дата обращения: 14.07.2025).
2. Шмелева Е. С. Фреймовые модели каламбура в интернет-мемах // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2020. № 9 (838). С. 173–183. EDN NDWWLR.
3. Якобсон Р. Язык в отношении к другим системам коммуникации // Р. Якобсон. Избранные работы / пер. с англ. Н. Н. Перцовой; сост. и ред. В. А. Звегинцев. М. : Прогресс, 1985.
4. Хрущева О. А. Роль мультимодальности при восприятии заимствований // Язык в эпоху цифровых трансформаций и развития искусственного интеллекта : сб. науч. ст. по итогам междунар. науч. конф., Минск, 23–24 окт. 2024 г. / отв. ред. Н. Е. Лаптева. Минск : МГЛУ, 2024. С. 506–511.
5. Кубрякова Е. С. О тексте и критериях его определения // Текст. Структура и семантика. Доклады VIII-ой Международной конференции (под ред. Е. И. Дибровой): сборник, Т.1. М. : СпортАкадемПресс, 2001. С. 72–81.

УДК 81`322

Шишкин Борис Андреевич

ассистент департамента лингвистики
Северо-Кавказский федеральный
университет
г. Ставрополь, Россия

Boris Shishkin

Assistant of the Department of Linguistics
North Caucasus Federal University
Stavropol, Russia
boris-shishkin.work@yandex.ru

ОПЫТ ТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ И КЛАСТЕРИЗАЦИИ В ИССЛЕДОВАНИИ ВИРТУАЛЬНОГО МЕДИАДИСКУРСА

В данной работе рассматривается контаминированный подход к проведению тематического моделирования и кластеризации текстов сетевого пространства с целью осуществления комплексного исследования виртуального медиадискурса. Базой исследования выступили отобранные методом сплошной выборки публикации из сетевого сообщества Северного Кавказа. Приводится опыт реализации автоматизированной систематизации эмпирического материала при помощи TF-IDF векторизации, кластеризации методом К-средних (k-means) и алгоритма тематического моделирования NMF. Выявляются сильные и слабые стороны данных подходов к анализу эмпирического материала. Отдельно отражается перспективность и эффективность дополнительного использования больших языковых моделей как инструмента автоматического именования сформированных кластеров и тематик. В заключении отмечается высокий потенциал данного подхода к комплексному рассмотрению состава сетевого пространства Северного Кавказа, что также служит основой для будущих исследований в области социолингвистики, лингвостатистики и иных направлений изучения виртуального медиадискурса.

Ключевые слова: тематическое моделирование; кластеризация; сетевое пространство; виртуальность; медиадискурс; К-средних; NMF.

EXPERIENCE IN TOPIC MODELING AND CLUSTERING IN RESEARCH OF VIRTUAL MEDIA DISCOURSE

This paper considers a contaminated approach to topic modeling and clustering of network space texts in order to carry out a comprehensive study of virtual media discourse. The study is based on posts from the North Caucasus online community selected using a continuous sample. The paper presents the experience of implementing automated systematization of empirical material using TF-IDF vectorization, K-means clustering, and the NMF topic modeling algorithm. The paper identifies the strengths and weaknesses of these approaches to analyzing empirical material. The paper indicates the prospects and effectiveness of additional use of large language models as a tool for automatic naming of formed clusters and topics. The conclusion notes the high potential of this approach to a comprehensive consideration of the composition of the North Caucasus network space, which can also be used as a basis for future research in the field of sociolinguistics, linguistic statistics and other areas of studying virtual media discourse.

Key words: topic modeling; cluster analysis; network space; virtual; media discourse; K-means; NMF.

В эпоху стремительно роста цифровых технологий виртуальная реальность, формируемая в сетевом пространстве, стала ключевым каналом коммуникации и источником знаний современного общества. Социальные сети, блоги, форумы, видеохостинги и другие платформы агрегируют колоссальный массив информации, продуцируемой различными пользователями интернета [1]. Данные материалы отражают не только субъективные позиции конкретных индивидов, но и формируют общее пространство дискурсивизации общественных настроений, культурных особенностей и социальных взаимодействий. Однако подобные свойства закономерно являются и препятствиями для исследования виртуального медиадискурса. Обширный объем данных, гетерогенность информации, полимодальность с трудом поддаются анализу при помощи ограниченных человеческих ресурсов.

Данная проблема особенно актуальна при анализе территорий с высокой динамикой коммуникации, где виртуальный медиадискурс отражает сложные социальные процессы, включая межличностные, межнациональные и религиозные отношения, традиции и политические события. Одним из таких субъектов можно назвать Северный Кавказ, являющийся как фронтальным, так и полиэтническим и поликонфессиональным регионом [2]. Потенциальный анализ виртуального медиадискурса подобного региона способен прояснить не только текущие общественные процессы, но и выявить потенциальные точки напряжения, конфликтогенные и синтонные тренды.

Преодолеть упомянутые выше препятствия и всесторонне проанализировать представленный медиадискурс возможно за счет внедрения алгоритмических процедур идентификации различного рода текстов [3]. В частности, продуктивно применение комплексного автоматизированного дискурсивномодусного подхода к исследованию интернет-ресурсов, одной из главных частей которого является тематическое моделирование и кластеризация сетевого пространства. Подобный подход позволяет сократить временные и трудовые затраты, обеспечивая высокую точность и интерпретируемость результатов, которые далее могут использоваться в иных интересах дискурсивных исследований. Таким образом, цель данной работы состоит в раскрытии сущности тематического моделирования и кластеризации как инструмента комплексного исследования виртуального медиадискурса.

Базой исследования выступил сформированный авторский корпус текстов, отобранных методом сплошной выборки из сообщества ВКонтакте «ЧП Кавказ Дагестан Чечня Ингушетия Осетия КБР» за период с 1 января 2021 г. по 1 июня 2025 г., что охватывает весь срок существования паблика на момент написания данной работы. Выбор данного сообщества обусловлен его широкой аудиторией (86 тыс. подписчиков), регулярностью освещения актуальных тем Северного Кавказа, активностью пользователей и репрезентативностью контента для различных этнокультур. Всего было собранно

3363 публикации, которые содержали новостные заметки, объявления, обсуждения и другие типы контента. Здесь можно заметить, что база исследования представляет собой обширный массив гетерогенных данных, что обуславливает целесообразность применения тематического моделирования и кластеризации в исследовании виртуального медиадискурса. Далее подробнее рассмотрим данные подходы к систематизации исследовательской информации.

Тематическое моделирование представляет собой «способ построения модели коллекции текстовых документов, которая определяет, к каким темам относится каждый из документов» [4]. Оно основано на предположении, что каждый текст представляет собой комбинацию нескольких тем, каждая из которых характеризуется определенным набором слов. При этом кластеризация фокусируется преимущественно на группировке текстов по принципу сходства. Однако здесь важно различать два подхода: жесткую кластеризацию, где каждый документ принадлежит ровно одному кластеру; и мягкую кластеризацию, когда документ может принадлежать нескольким кластерам одновременно, что сближает её с представлением документов в тематическом моделировании.

В данной работе процесс тематического моделирования и кластеризации основывался на применении алгоритма неотрицательного матричного разложения (англ. *NMF*) и метода К-средних (англ. *k-means*) к текстам, прошедши предварительную векторизацию посредством расчета *TF-IDF* меры.

Отдельно поясним, *TF-IDF* (англ. *Term Frequency-Inverse Document Frequency*) – это метод векторизации текстов, который оценивает значимость слов в документах относительно всего корпуса, учитывая частоту слова в тексте (*TF*) и обратную частоту его появления в корпусе (*IDF*). Кластеризация методом К-средних группирует тексты на основе их векторного представления, выявляя тематические кластеры. *NMF*, в свою очередь, разлагает матрицу *TF-IDF* на две неотрицательные матрицы, представляющие темы и их веса в документах, а также слова и их веса в темах, что позволяет выявить мягкие тематические структуры с высокой интерпретируемостью [5, р. 86]. Представленный подход также дополняется использованием больших языковых моделей для автоматического именования тем/кластеров, что повышает точность и скорость анализа результатов.

Стоит отметить, что реализация упомянутых методов осуществлялась на основе программ, написанных на языке *Python* с использованием: специализированных библиотек для машинного обучения, а именно *scikit-learn*; *API* ВКонтакте для сбора первичного эмпирического материала; *API DeepSeek* с целью осуществления прямых запросов к большой языковой модели для автоматического именования тем/кластеров.

В результате выбор подобного подхода к исследованию способствует эффективной автоматизации процессов анализа обширных и специфических

текстов социальных сетей. Общий алгоритм проведения тематического моделирования и кластеризации виртуального медиадискурса можно представить в виде схемы, визуализирующей данный процесс от этапа сбора данных до итогового анализа полученных результатов (Рисунок 1).

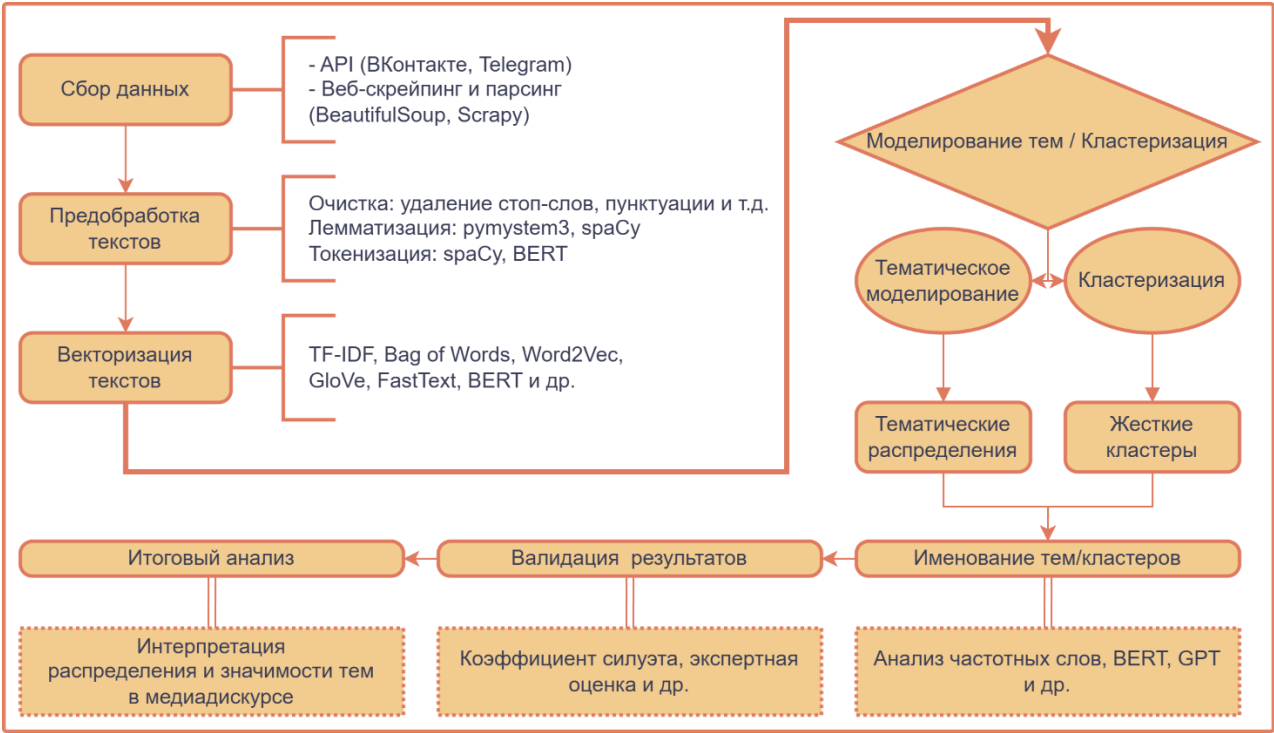


Рис 1. Алгоритм проведения тематического моделирования и кластеризации виртуального медиадискурса

В рамках исследования на основе метода К-средних с дополнительным применением коэффициента силуэта было выявлено 8 кластеров, соответствующих основным темам, представленным в отобранных публикациях социальных сетей. При помощи алгоритма *NMF* также было выделено аналогичное количество тематик анализируемого эмпирического материала. Далее рассмотрим конкретные выявленные кластеры и их наполнение:

- 1) ЧП и происшествия (310 записей). Ключевые слова по *TF-IDF* мере: пожар, махачкала, писать, мы, пострадать, гореть;
- 2) Дети и семья (348 записей). Ключевые слова по *TF-IDF* мере: ребёнок, девочка, мальчик, дом, летний, год;
- 3) Реклама чата (33 записи). Ключевые слова по *TF-IDF* мере: репост, чат, телеграм, chat, заходить, спать;
- 4) Криминальные происшествия (163 записи). Ключевые слова по *TF-IDF* мере: девушка, её, парень, который, летний, махачкала;
- 5) Разные инциденты (646 записей). Ключевые слова по *TF-IDF* мере: мужчина, летний, полиция, полицейский, парень, сотрудник;

6) Новости Кавказа (1257 записей). Ключевые слова по *TF-IDF* мере: дагестан, видео, район, наш, махачкала, свой;

7) Дорожные происшествия (309 записей). Ключевые слова по *TF-IDF* мере: дтп, погибнуть, водитель, авария, пострадать, автомобиль;

8) Преступления и инциденты (295 записей). Ключевые слова по *TF-IDF* мере: летний, уголовный, дело, житель, подозревать, мужчина, задержать.

Аналогичным образом представим обнаруженные тематики алгоритмом *NMF*:

1) Преступления и происшествия (491 запись). Ключевые слова: летний, житель, мужчина, задержать, дело, уголовный, подозревать, убийство, возбудить;

2) ДТП и аварии (423 записи). Ключевые слова: дтп, водитель, погибнуть, автомобиль, авария, результат, пассажир, происшествие, пострадать;

3) Приглашения в чаты (53 записи). Ключевые слова: репост, чат, телеграм, chat, заходить, спать, телеграм, беседа, максимальный;

4) Криминальные происшествия (1040 записей). Ключевые слова: девушка, парень, видео, деньга, её, день, человек, женщина, полиция;

5) Чрезвычайные ситуации (295 записей). Ключевые слова: пожар, пострадать, произойти, информация, гореть, дом, взрыв, пожарный, мчс;

6) Криминал и происшествия (276 записей). Ключевые слова: махачкала, писать, акушинский, сегодня, проспект, драка, хасавюрт, кизилюрт, вчера;

7) События в регионах (416 записей). Ключевые слова: дагестан, район, село, уточняться, писать, кизилюртовский, цунтинский, казбековский, утонуть;

8) Трагедии и дети (367 записей). Ключевые слова: ребёнок, мальчик, девочка, мать, дом, родитель, больница, мама, отец.

Как можно заметить, обнаруженные кластеры и темы в целом совпадают друг с другом, отражая схожие семантические характеристики анализируемых публикаций. Однако алгоритм *NMF* закономерно показал лучшие результаты в распределении записей по конкретным темам. Например, короткая рекламная запись: *НАША БЕСЕДА В ТЕЛЕГРАММЕ* [6], которая относится к кластеру *Новости Кавказа*, была помечена темой *Приглашения в чаты*. В целом различия в распределении публикаций на темы и кластеры возможно проследить на представленной ниже тепловой карте (Рисунок 2).

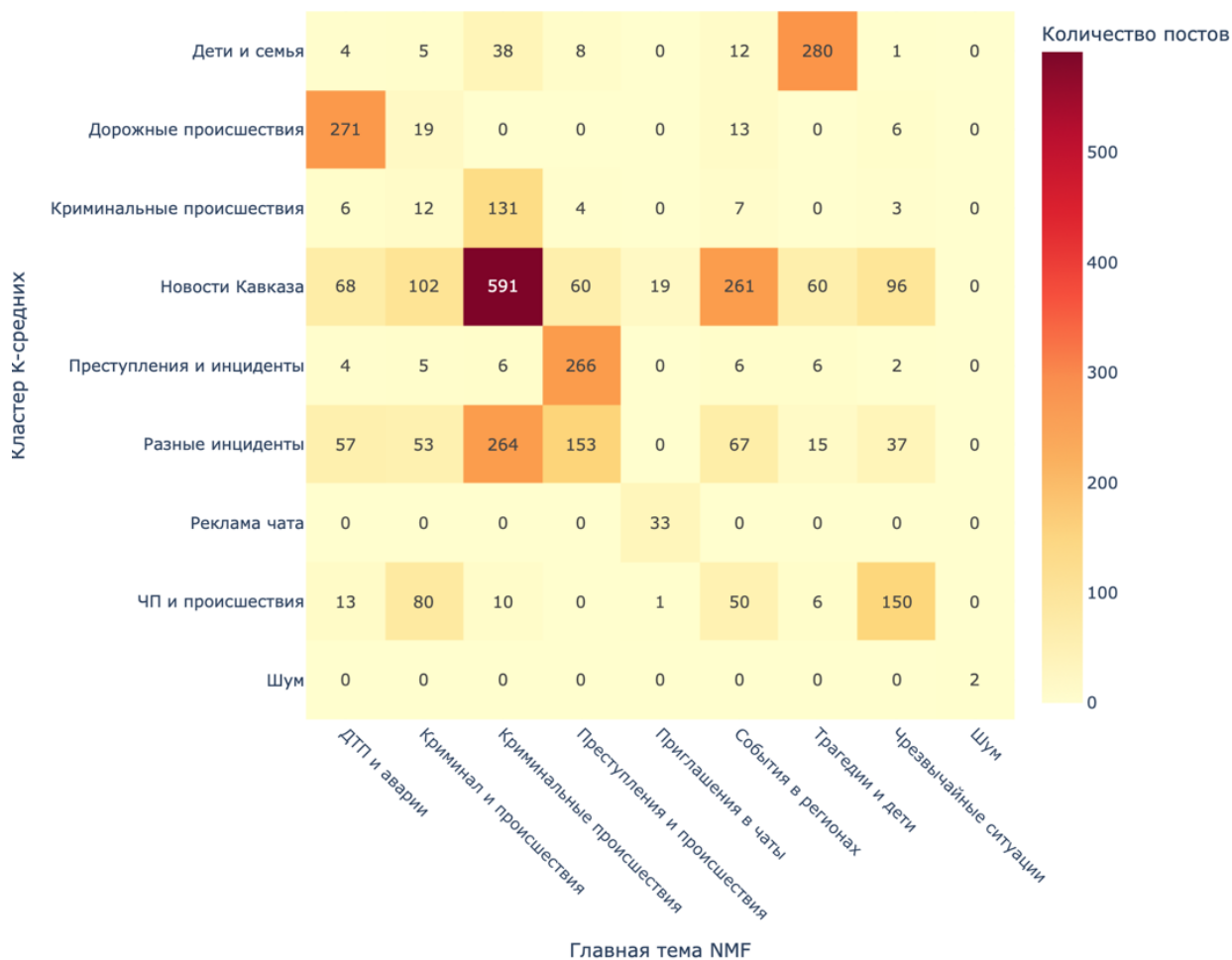


Рис 2. Распределение постов по кластерам К-средних и главным тематикам NMF

Алгоритм NMF, как уже было упомянуто ранее, способен также выделять и дополнительные темы в конкретных текстах. Здесь в пример можно привести следующую публикацию:

Поисково-спасательные работы идут в Бежтинском участке, где 11 мая пропал мальчик 2021 года рождения, сообщает МЧС по Дагестану. Предполагают, что ребёнок мог упасть с моста в реку. Нашли его велосипед, висящий на опоре моста. В поисках участвуют 21 спасатель, в том числе кинолог с собакой, а также 50 сотрудников МВД и местные жители [7].

Данная запись ожидаемо была отнесена к кластеру 2 *Дети и семья* с главной темой 8 *Трагедии и дети*. Однако в этой публикации также были обнаружены элементы и других тем, а именно 1 *Преступления и происшествия*, 7 *События в регионах*, 5 *Чрезвычайные ситуации*.

Отдельно следует отметить, и метод К-средних, и алгоритм NMF имеют как свои слабые, так и сильные стороны. К-средних достаточно прост в реализации, имеет высокую скорость и предоставляет четкое разделение текстов на группы. Однако данный подход игнорирует семантические связи,

а также чувствителен к шуму. NMF более гибок, предоставляет мягкое тематическое распределение с высокой интерпретируемостью, но сложнее в настройке и также чувствителен к шуму. В итоге выбор между этими методами сводится к поставленным задачам исследования, а их сочетание способствует комплексному анализу виртуального медиадискурса.

Таким образом, проведенное исследование демонстрирует эффективность тематического моделирования и кластеризации при анализе обширных массивов данных виртуального медиадискурса. При этом использование больших языковых моделей для автоматического именования кластеров и тематик повышали интерпретируемость результатов. В целом данный подход способствовал комплексному рассмотрению состава сетевого пространства Северного Кавказа, что также может послужить основой для будущих исследований в области социолингвистики, лингвостатистики и иных направлений.

ЛИТЕРАТУРА

1. Шишкин Б. А. Вариативность нарратива гармонизации сетевой коммуникации в медиадискурсе [Электронный ресурс] // Актуальные проблемы филологии и педагогической лингвистики. 2024. № 3. С. 135–150. DOI: <https://doi.org/10.29025/2079-6021-2024-3-135-150>.
2. Социокультурные основы Российской идентичности молодежи Северного Кавказа: ценностные трансформации и сетевые репрезентации / А. М. Ерохин, Е. А. Авдеев, С. Н. Бредихин, С. М. Воробьев, Б. А. Шишкин. Ставрополь : ООО «Бюро новостей», 2024. 324 с.
3. Каменский М. В., Бредихин С. Н. Алгоритмические процедуры идентификации рекламных текстов в дискурсивном пространстве средств массовой информации // Вестник Волгоградского государственного университета. Серия 2: Языкознание. 2025. Т. 24, № 1. С. 64–78. DOI: <https://doi.org/10.15688/jvolsu2.2025.1.6>.
4. Коршунов А., Гомзин А. Тематическое моделирование текстов на естественном языке // Труды Института системного программирования РАН. 2012. Т. 23. С. 215–44.
5. Aggarwal C.C., Zhai C. A Survey of Text Clustering Algorithms [Electronic resource] // Mining Text Data. Boston : Springer, 2012. P 77–128. DOI: https://doi.org/10.1007/978-1-4614-3223-4_4.
6. Ночной чат // ЧП Кавказ Дагестан Чечня Ингушетия Осетия КБР. URL: https://vk.com/wall-201560969_256536 (дата обращения: 20.06.2025).
7. Поисково-спасательные работы идут в Бежтинском участке // ЧП Кавказ Дагестан Чечня Ингушетия Осетия КБР. URL: https://vk.com/wall-201560969_338817 (дата обращения: 20.06.2025).

УДК 811.133.1

Яскевич Денис Михайлович
аспирант
Московский педагогический
государственный университет
г. Москва, Россия

Denis Iaskevich
Phd Student
Moscow State Pedagogical University
Moscow, Russia
denyasden@yandex.ru

ОСОБЕННОСТИ СЕТЕВОГО ДИСКУРСА ВО ФРАНКОЯЗЫЧНОМ ИНТЕРНЕТ-ПРОСТРАНСТВЕ

В статье анализируется сетевой дискурс франкоязычных политематических сообществ на платформе Discord, одной из ведущих онлайн-платформ для виртуального общения. Объектом исследования выступают письменные фрагменты пользовательских диалогов и полилогов, позволяющие выявить лингвистические особенности современной интернет-коммуникации. На материале диалогических и полилогических фрагментов общения изучаются структурно-семантические и прагматические особенности цифрового дискурса, включая аграмматизм, экономию языковых усилий, сочетание литературного и разговорного стилей, гибридизацию, заимствования и влияние фактора анонимности на речевое поведение. Исследование показывает, что пользователи в стремлении к экспрессивности и эффективности взаимодействия активно нарушают нормы орфографии и синтаксиса, используют нестандартные графические средства и креативные сокращения, что формирует особую языковую реальность. Платформа Discord предоставляет уникальную эмпирическую базу для наблюдения за тем, как формируется виртуальная разновидность французского языка, адаптированная под условия спонтанного цифрового взаимодействия. Работа вносит вклад в описание механизмов сетевого словотворчества и актуализирует вопрос о трансформации языковых норм в условиях онлайн-коммуникации.

Ключевые слова: сетевой дискурс; онлайн-платформа; виртуальная личность; гибридизация; цифровая коммуникация.

SPECIFIC FEATURES OF NETWORK DISCOURSE IN THE FRANCOPHONE ONLINE SPACE

This article examines the network discourse of French-speaking thematically diverse communities on the Discord platform, one of the leading online environments for virtual communication. The study focuses on written fragments of user dialogues and polylogues to identify the linguistic features of contemporary internet interaction. Based on dialogic and polylogic communication excerpts, the research analyzes the structural-semantic and pragmatic characteristics of digital discourse, including agrammatism, linguistic economy, the blending of literary and colloquial styles, hybridization, lexical borrowing, and the influence of anonymity on speech behavior. The findings show that users, in their pursuit of expressiveness and communication efficiency, often violate orthographic and syntactic norms, make use of non-standard graphic elements and creative abbreviations, thereby shaping a distinct linguistic reality. The Discord platform offers a unique empirical base for observing the emergence of a virtual variety of the French language, adapted to the conditions of spontaneous digital interaction. The study contributes to the description of mechanisms of online word formation and raises the issue of transforming linguistic norms in the context of online communication.

Key words: network discourse; online platform; virtual identity; hybridization; digital communication.

С развитием технологий онлайн-среда стала не только источником информации и платформой для общения, образования, работы, развлечений, но и породила новые виртуальные среды и сообщества. Развитие интернет-платформ, помимо своих практических целей, способствовало появлению совершенно новых отношений между людьми в обществе. В частности, появилась особая форма коммуникации – интернет-коммуникация, а следовательно, новая культурная и языковая среда.

В данной статье рассматриваются фрагменты сетевого дискурса франкоязычных серверов политематических сообществ кроссплатформенного мессенджера Discord, который за период с 2020 по 2024 г. стал одним из наиболее популярных онлайн-платформ, используемых франкоязычными пользователями. Так, в 2020 г. количество уникальных ежемесячных посетителей анализируемой нами платформы составило 7,9 миллионов, что стало на 50 % больше по сравнению с 2019 г. К 2023 г. Статистические службы сообщили о 2,2 миллионах ежедневных пользователей [1]. По данным 2024 г. Франция вошла в топ-5 стран по количеству генерации трафика в браузерной версии (3,43 % от общего количества пользователей) [2]. Кроме того, во Франции имеется больше серверов, количество участников которых превышает 30 человек, что превышает средние показатели по миру. При этом чаще пользователи кроссплатформенных систем мгновенного обмена сообщениями во Франции чаще подключаются с мобильных устройств, что также является рекордным показателем.

К началу 2024 г. года платформа Discord зафиксировала 614 миллионов зарегистрированных пользователей, из которых ежедневно платформу посещает около 29 миллионов человек [3]. Благодаря техническим ресурсам и открытому доступу к серверам имеется возможность анализировать гипертекстовые единства, отражающие наиболее значимые особенности и характеристики интернет-общения, в том числе способы авторизации франкоговорящих пользователей в сетевом дискурсе.

Исследование диалогических и полилогических единств обеспечивает наблюдение за письменной речью пользователей в естественных условиях, где нормы сетевого общения формируются спонтанно и подвергаются минимальной редакции. Такое разнообразие исследуемых площадок позволяет регистрировать как общепринятые аббревиатуры и цифровые замены, так и локальные креативные практики, что, в свою очередь, создаёт необходимую эмпирическую базу для дальнейшего описания механизмов сетевого словотворчества и для сопоставительного анализа устойчивых и вариативных тенденций фонетико-графической экономии в современном франкоязычном виртуальном дискурсе.

Коммуникативное поведение участников серверов включает следующие тенденции: тенденция к аграмматизму, тенденция к экономии языковых усилий и её нарушение с целью добавления в высказывание аффективной составляющей, сочетание литературного и разговорного языковых стилей,

использование гибридных лексических единиц и заимствований, переход к неформальному обращению на «ты» и использованию обценной лексики из-за фактора анонимности.

1. Тенденция к аграмматизму

Стабильное нарушение орфографических, синтаксических и пунктуационных норм языка может рассматриваться в качестве одной из основополагающих характеристик интернет-дискурса и коммуникативных ситуаций в виртуальной среде в частности, что связано со стремлением пользователей в ускоренном порядке передать содержательную сторону сообщения, не обращая внимания на возможные ошибки и опечатки, которые не мешают восприятию информации [4, с. 128–129]. Так, в полилоге на тему спорта зафиксирован ряд характерных для сетевого общения грамматических и орфографических отклонений, которые демонстрируют тенденцию к аграмматизму в условиях спонтанного цифрового взаимодействия. К ним можно отнести отсутствие флексии «s» у глагола настоящего времени второго лица единственного числа ‘tu t’entraîne’, использование притяжательного прилагательного sa вместо указательного местоимения ‘ça’, употребление неверной флексии для вспомогательного глагола avoir в imparfait третьего лица единственного числа, лишнее согласование ‘sa avais durée’, подмена инфинитива формой причастия прошедшего времени ‘tu vas vite te découragé’, и наоборот ‘j’commencer a gagner’, а также отсутствие диакритических знаков в лексемах и предлогах ‘a jeun’, ‘a gagner’, ‘petit dej’, ‘apres’ (Таблица 1).

Таблица 1

Тема беседы: спорт, часть 1

Пример	Перевод
212 : bh c’est galère surtout si en même temps tu t’entraîne tu verras ton poids il bouge pas Tabarnus : cardio a jeun 212 : bh c’est galère surtout si en même temps tu t’entraîne tu verras ton poids il bouge pas Tabarnus : cardio a jeun 212 : tu vas vite te découragé Tabarnus : et gros petit dej apres ishralam : Après le poids qui stagne c normal , quand j’faisais de la natation mon poids s’était stagné pendant 1–2 semaines pour que mon corps s’habitue aux changements Ça peut durer sur 1 mois le temps que ton corps s’habitue à ton nouveau rythme Tabarnus : oui ben moi il s’habitue depuis 6 mois lol 212 : nah moi sa avais durée 1 semaine	212: ох, это мучение особенно если в это время ты тренируешься ты увидишь, твой вес, он не меняется Tabarnus : кардио натошак 212: ох, это мучение особенно если в это время ты трени-руешься ты увидишь, твой вес, он не двигается Tabarnus: кардио на голодный желудок 212: ты скоро потеряешь мотивацию Tabarnus: и плотный завтрак после ishralam: После этого застой веса – нор- мальное явление, когда я занимался пла- ванием, мой вес не менялся в течение 1–2 недель, чтобы организм привык к изменениям Это может длиться месяц, пока организм привыкает к новому ритму Tabarnus: да, ну я привыкал к нему 6 месяцев лол 212: не, у меня это длилось неделю

ishralam : En vrai oui , tjr courir le ventre vide j'ai fait un semi-marathon en ayant mangé juste avant , failli vomir mes mooorts 212: puis jcommencer a gagner	ishralam: да, действительно, я всегда бегая на голодный желудок, я пробежал полумарафон, поев перед этим, и меня чуть не стошнило 212: потом я начал набирать

Несмотря на наличие вышеописанных отклонений от нормативных языковых стандартов, семантическая эффективность сообщений сохраняется, поскольку описанные типы ошибок не критичны для восприятия письменного текста и, как правило, не препятствуют интерпретации высказываний участниками дискурса. В подобных ситуациях виртуальная коммуникативная беседа продолжается беспрепятственно, то есть без дополнительных запросов на уточнение. Тем не менее при необходимости самокоррекции уже опубликованного сообщения пользователь в следующем сообщении может указать корректную форму после символа «*», что привлечёт внимание адресатов к внесённому уточнению (Таблица 2).

Таблица 2

Тема беседы: спорт, часть 2

Пример	Перевод
Tabarnus: maintenant je mets forcement un t shirt quand jenage dehors bon de toute facon he bage okus la ishralam: Ah gars surtout jsp.comment c la météo en Asie du sud mais ça m'a l'air d'être piquant Tabarnus: *je nage	Tabarnus: теперь мне приходится надевать футболку, когда я плаваю не в помещении что ж, в любом случае (неразборчиво) здесь ishralam: Ах, ребята, я не знаю, какая погода в Южной Азии, но похоже, что она будет жаркой Tabarnus: *я плаваю

2. Тенденция к экономии языковых усилий

Тенденция к экономии усилий при использовании языковых средств является преобладающей в коммуникативных ситуациях, где собеседникам необходимо быстро обменяться информацией. Так, в чатах и на форумах политематических серверов Discord франкоговорящие пользователи активно прибегают к сокращениям, усечённым лексемам и аббревиатурам, часть из которых также применяются во время SMS-общения. Одним из наиболее популярных методов сокращения лексем становится «удаление» гласных из слова, после чего полученный термин идентифицируется по оставшимся согласным. Так, например, лексема *beaucoup*, состоящая из восьми букв, сокращается до трёх смыслоразличительных согласных – *бср* – и активно используется как в SMS-языке, так и при виртуальных переписках для экономии времени. На сервере *Chicken Wings* зарегистрировано более 1 тыс. употреблений данного сокращения в сообщениях участников сервера.

Исходя из переписок франкоязычных пользователей в чатах и на форумах, можно прийти к выводу, что участники бесед предпочитают использовать подобные сокращения для экономии времени, затрачиваемом на переписку, поскольку нажать на клавиатуре три клавиши для написания 'bsr' вместо восьми представляется более быстрым и удобным. Подобные аббревиатуры и сокращения довольно известны широкой аудитории и не призваны кодировать смысл слова, а лишь сократить количество нажатий на клавиатуре при их наборе:

(1) а. **Pk tu dit chui p3 ?** (пользователь Mo.Salah) (Почему ты говоришь, что я в рейтинге топ-3?)

б. **Askip** je suis inactive (пользователь Melo) (Видимо, я не онлайн)

с. **Jsp** je suis enjoué pour rien mais depuis tous a l'heure j'enchaîne entre ça et être triste... Je sais **mm** pas comment je suis sencé ressentir ça... (пользователь Zana_love) (Я не знаю, я веселюсь по пустякам, но с тех пор я постоянно переключаюсь между радостью и грустью... Я не знаю, что я должен чувствовать...)

д. Ouais on gros on travail la construction d'un discours en public, que se soit les méthode sur le vocabulaire les figure de style à utiliser afin d'affiner un argumentaire percutant tout comme sa position, l'encrage au sol, gestuel etc en soi c pas mal mais là c'est juste chiant... (пользователь Sensei) (Да, в основном мы работаем над построением речи на публике, будь то словарный запас или фигуры речи, которые нужно использовать, чтобы отточить убедительную аргументацию, или будь то поза, положение стоя, язык тела и т.д. Само по себе это неплохо, но сейчас это просто скучно...).

Лексический уровень включает использование нехарактерных для устной разговорной речи аббревиатур и сокращений, которые появились с целью сокращения времени написания и, как следствие, иллюстрирующие актуальность тенденции к экономии языковых средств. Как можно заметить в примере (3) д, 'on travail' используется вместо 'on travaille', 'que se soit' вместо 'que ce soit', 'les méthode' и 'les figure' вместо 'les méthodes' и 'les figures', в сетевом дискурсе стремление сократить время написания сообщения может приводить к аграмматизму и совершению опечаток, что воспринимается пользователями в основном нейтрально.

Активное использование аббревиатур и сокращений наиболее ярко демонстрирует стремление пользователей к экономии языковых средств и, как следствие, к экономии времени, затраченном на переписку. Так, поздороваться и попрощаться можно с помощью сокращений 'bjr' (bonjour) и 'bsr' (bonsoir), 'с-а-д' используется вместо выражения 'с'est-à-dire', а чтобы попросить прощения у собеседника, можно использовать сокращение 'dsl' (désolée). Вместо лексемы 'aujourd'hui' используется 'ajd', а слово 'toujours' превращается в 'tjr'. Также удобство данного способа заключается в том, что пользователи могут более выразительно передать необходимую информацию или передать эмоции, при этом выбрав сокращённый способ: 'ptdrrrr, PTDR,

pdr, mdr, mdr' (pété de rire, mort de rire – умираю со смеху). Для экономии времени и места используются также эллиптические конструкции, при этом автор стремится выразить максимальное количество информации с помощью минимального набора языковых средств: 'après c' вместо 'après ça', 'puis c un passage' вместо 'puis c'est un passage'. Как можно заметить, в эллиптических конструкциях с указательным местоимением 'ça' также пропадает диакритический знак [5, с. 523].

2 (a) *Нарушение тенденции к экономии языковых усилий с целью добавления в высказывание аффективной составляющей*

Целесообразно подчеркнуть, что добавление аффективной составляющей в реплики может нарушать принцип экономии языковых средств, поскольку оно предусматривает использование дополнительных графических и пунктуационных средств для создания акцента на определённой аффективной окраске, транслируемой собеседникам в зависимости от контекста беседы. Так, для большей эмоциональной насыщенности пользователи могут записывать слово или выражение полностью заглавными буквами, выделять основную мысль полужирным шрифтом и / или цветом:

(2) a. НАНАНАНАНА С EXTREMMENT DROLE (пользователь tweakerfromvbeach) (XAXAXAXAXA ЭТО БЕЗУМНО СМЕШНО)

b. J'ATTEND LA SUITE (пользователь Le Friezaiolo) (Я ЖДУ ПРОДОЛЖЕНИЕ)

c. Tu fais **exprès** de m'ignorer encore je te spam mp (second degré officiellement) (пользователь GDP/Zana) (Ты опять **специально** игнорируешь меня, я спамлю тебя в личных сообщениях (вторая степень официально))

Одним из способов придания высказыванию экспрессивности становится многократное повторение одного или нескольких знаков препинания, а также гласных или согласных букв слова для имитации некоторых супrasegmentных языковых явлений:

(3) a. Ya qu*oiiii*i dans 72 **hhhhh** (пользователь ♡°•○Nova°•°♡) (Чтооо там через 72 часаааа)

b. Bahahah c'est toi le petit cochon qui enchainee JE SAVAIS !!! (пользователь ♡°•○Nova°•°♡) (Ахахаха это ты маленький поросёнок в цепях Я ЗНАЛ !!!)

Коммуникативное поведение некоторых отдельных пользователей может включать в себя применение определённых стратегий, отражающих индивидуальность участника сервера, становясь узнаваемой манерой общения. Так, к примеру, речевое поведение пользователя ♡°•○Nova°•°♡ в анализируемых репликах включает в себя большое количество лексем с множественными повторами гласной или согласной буквы: pourquoiiii, bruhh, pffff, uiiii, tsssss, ahahha, chilll, bah chuttt, ptinnn, mdrrrrrrr. В подобном случае целесообразно говорить о графической имитации интонации, чтобы создать впечатление, что собеседник произносит слова «нараспев», выражая нетерпение (dittt laaa), раздражение (bah chuttt, ptinnn, tsssss), отвращение (bruhh) или радость (uiiii,

ahahha). Данный способ имитации интонационных супrasegmentных явлений привносит в реплики ♡°•○Nova°•°♡ различную аффективную окраску, экспрессивность.

3. Сочетание литературного и разговорного языковых стилей

Коммуникативное поведение франкоязычных пользователей в интернет-пространстве реализует игровое начало виртуального общения и содержит как черты литературного языка, так и элементы его разговорного варианта. Помимо использования лексических единиц, характерных для разговорного уровня языка, характерными чертами поведения франкоговорящих пользователей во время виртуального общения в сети на пунктуационном уровне являются отсутствие точки в конце предложения и вместе с тем активное использование одного или нескольких вопросительных или восклицательных знаков для добавления аффективной коннотации. На орфографическом уровне можно отметить несоблюдение правила написания начала предложения, имён или названий стран с заглавной буквы, возможное отсутствие расстановки диакритических знаков, фонематическое написание слов:

(4) a. pour l'instant il n'est disponible qu'en anglais, j'avoue je ne sais pas comment ça se passe pour ce genre de films, s'ils vont un jour connaître une sortie en **france** ? (пользователь supr_nathan) (на данный момент он доступен только на английском языке, признаю, я не знаю, что бывает с такими фильмами, выйдут ли они когда-нибудь во Франции?)

b. c'est assez **recent** en **europe** (пользователь GaB) (В Европе это появилось совсем недавно)

c. dans mon école on devait faire un voyage en **angleterre** en famille partagé ou jsp comment ça s'appelle (пользователь les_3_4_5) (В моей школе мы должны были отправиться в поездку в Англию с проживанием в семье, или как это называется, я не знаю)

4. Использование гибридных лексических единиц

Нередко интернет-пользователи во время виртуальной коммуникации прибегают к использованию гибридизированных лексических единиц. Гибридизация как лингвистический феномен представляет собой универсальный процесс, способствующий формированию новых лексических единиц за счёт объединения элементов различной этимологии, стилевой или знаковой природы. Особое значение процесс гибридизации приобретает в рамках виртуального дискурса, где за счёт взаимодействия разнородных языковых ресурсов формируются неологизмы с расширенным семантическим и функциональным потенциалом. Гибридные единицы, возникающие как результат деривации, заимствования и языковой интерференции, отражают динамику современной лексикосистемы и свидетельствуют о тенденции к интеграции и трансформации языковых структур в условиях цифровой коммуникации. Гибридизация представляется одним из ключевых факторов неологизации и развития лексической системы в условиях сетевого дискурса.

В переписках пользователей онлайн-платформы политематических сообществ Discord можно встретить следующие гибриды, в которых произошла

замена буквосочетаний, соответствующих фонеме [ɛ̃], на цифру «один» (un): 1posibl / 1possible (impossible), je / tu / il / elle / on vi1 (viens / vient), которое чаще всего можно встретить в форме вопроса tv1? (tu viens ?), m1 (main), m1tnan (maintenant), l1di (lindi), k11 (câlin), a12c4 (à un de ces quatre), qq1 / kk1 (quelqu'un), ri1 (rien), gf1 (j'ai faim), pt1 (putain), ter1 (terrain), europé1 (européen), а также bi1 (bien) и его производные (ab1to – à bientôt, b1sur – bien sûr):

- (5) a. vi1 la (пользователь Yu) (приходи сюда)
- b. vi1 on fait notre mariage discord ici et IRL tq1 je m'arrangerai (пользователь Right There) (приходи, у нас будет свадьба здесь, в Discord, и в реальной жизни (IRL – от англ. in real life), не волнуйся, я всё устрою)
- c. Marine tv1?? on commence dans 10 m (пользователь Malico) (ты идёшь?? мы начинаем через 10 минут)
- d. vi1 v0c (пользователь Teebowzz) (переходи в голосовой чат)
- e. Bonjour tout le monde si qq1 peut me rendre un ptit service svp (пользователь matthew_gem) (Всем добрый день, может ли кто-нибудь оказать мне небольшую услугу, пожалуйста)

Использование гибридизированных лексем, состоящих одновременно из букв и цифровых символов, может стать причиной орфографических искажений: Pt1 je paierai cher pour un Taco bell (пользователь Sulla), j'ai donner 20\$ à kk1 pour un investissement jsp quoi (пользователь Teebowzz).

5. Использование заимствований

Для молодого франкоязычного поколения довольно привычно использование англицизмов во время переписки, что помогает им продемонстрировать свою индивидуальность и почувствовать себя частью международного сообщества. В анализируемых фрагментах встречались следующие заимствования из английского и арабского языков: 'Lets go, fuck it, how, gg, lol, sooo, уур / уер, wtf, nah, tips, noob, un like, un hashtag, un loser, wsh, sbah'.

При анализе фрагментов сетевого дискурса было также зафиксировано, что участники политематических серверов исследуемой платформы, среди которых большую часть составляет молодая аудитория [6], в процессе виртуальной коммуникации могут сделать выбор как в пользу заимствованной лексемы, так и в пользу термина из родного языка. Так, при поиске количества употреблений лексемы smart (умный, умно) общий поисковик чата показывает 741 результат (bah il est un peu pas smart, faut penser smart mec, franchement elle etait ultra smart и т.д.), в то время как термин intelligent в подобных контекстах использовали в 823 случаях (T vrmt intelligent, Il doit être big intelligent, Nan vrm lui il est intelligent и т.д.).

6. Влияние фактора анонимности на ход языковых практик

Возможность сохранения анонимности для пользователей может оказаться фактором, обуславливающим использование местоимения «ты» по отношению к другим пользователям сервера, а также использование обценной лексики, оскорбительных выражений и неуважительное поведение с коммуникантами в конфликтных ситуациях. Поскольку во французском языке

существует местоимение второго лица множественного числа «вы», не совпадающее с местоимением второго лица единственного числа «ты» и являющееся его более вежливым аналогом, то употребление местоимения «ты» к незнакомому человеку может быть расценено как знак неуважения. Однако в сетевом дискурсе демонстрируется склонность к использованию местоимения «ты» и в ряде проанализированных коммуникативных ситуаций не обладает намерением унижить собеседника. Его использование также обуславливается фактором анонимности, которое, с одной стороны, не раскрывает личности пользователя, а с другой – не позволяет узнать личность собеседника, следовательно, определение его возрастной категории и статуса в социуме становится невозможным:

(6) а. **T** as quel age toi ? (пользователь Trizzy) (А тебе сколько лет?)

б. @Trizzy **toi** qui de mémoire est sur Toulouse, y aurait **tu** des endroits à me recommander pour acheter des DVD le moins cher possible ? Neufs ou d'occasion peu importe. (пользователь Cococinelle) (@Триззи, ты, насколько я помню, живёшь в Тулузе, не мог бы ты порекомендовать места, где я могу купить DVD-диски как можно дешевле? Новые или подержанные, неважно.)

с. ce que **tu** peu aussi faire, c'est de faire **ta** propre video (пользователь custos) (что ты можешь ещё сделать, так это снять собственное видео)

Тем не менее бывают ситуации, когда демократичный характер коммуникации превращается в пренебрежительный. На рассматриваемых нами серверах политематических виртуальных франкоязычных сообществ в разделе личной информации о пользователе могут встретиться только те сведения, которые пользователь сам решит добавить в раздел «обо мне», остальная информация не является персонализированной, следовательно, не приводит к конкретному человеку или личному профилю этого человека на других виртуальных платформах.

В ряде наблюдаемых ситуаций взаимодействие между участниками виртуальной коммуникации на онлайн-платформе происходит на одном уровне, без какого-либо ориентира на социальную иерархическую структуру, принятую в обществе, что, в свою очередь, оказывает существенное влияние на подбор языковых средств. К тому же, выбор дистанции, которую следует держать с собеседником в виртуальном пространстве, зависит от «степени видимости» личных и индивидуальных данных, которыми делится пользователь в своём аккаунте [7, р. 154]. Так, пользователи с меньшим количеством личной информации в своём аккаунте намного чаще пренебрегают иерархической структурой и прибегают к использованию местоимения второго лица единственного числа *tu* вместо местоимения множественного числа *vous*, пренебрежительно коннотированных номинантов и выражений, разговорных конструкций, относящихся к стилю *familier*, а в случае проявления вербализированной агрессии могут использовать уничижительные, оскорбительные языковые выражения. Вместе с тем, одной из неотъемлемых составляющих интернет-коммуникации является возможность свободно выражать свои

мысли, которая иногда граничит с вседозволенностью. Следствием данного явления становится эскалация негативных оценочных суждений в сторону других пользователей виртуального пространства (Таблица 3).

Таблица 3

Тема беседы: приветствие пользователей в чате

Пример	Перевод
OZEL: wsh les marseillais Alastor: c'est quoi ces conv de segpa patachong: ferme ta chatte	OZEL: привет, жители Марселя, как дела Alastor: что это за беседа в коррекционном классе patachong: закрой свой рот

В данном полилоге наблюдается алжирско-арабское заимствование ‘wsh’, означающее приветствие и вопрос о том, как поживают собеседники. Ответная реплика содержит сравнение беседы с той, которая может звучать среди учеников коррекционных классов (Segpa: Section d'enseignement général et professionnel adapté – специальный коррекционный класс общего и профессионального образования), что моментально расценивается пользователями как оскорбление, в связи с чем patachong отвечает пользователю Alastor в грубой форме.

Использование ненормативной лексики демонстрирует устойчивую рас пространенность в коммуникативных практиках сетевого франкоязычного пространства. В частности, франкоязычные пользователи регулярно прибегают к использованию лексемы ‘con’, которая, функционируя в качестве оскорбительного обращения и средства выражения пренебрежения/презрения, позволяет производить оценочную категоризацию как поведения окружающих субъектов, так и рефлексивно оценивать собственные действия:

(7) a. mais ce **con** a déboulé en pleine nuit sur la route (пользователь Manus)
(но этот идиот выскочил на дорогу посреди ночи)

b. oue с **con** lol (пользователь Manus) (да, это глупо, лол)

Таким образом, было установлено, что пользователи транслируют субъективную картину мира, что отражается в языке. Стремление донести до собеседника сообщение в сжатой форме, полностью сохранив смысловую и эмоциональную нагрузку высказывания, порождает структурно-семантические изменения в языке, которые отражаются на фонетико-графическом, морфолого-синтаксическом, лексико-пиктографическом языковых уровнях. Появляется совершенно новый язык виртуальной коммуникации, обладающий особенностями и характеристиками, отличными от тех, которые присущи его нормативному варианту.

ЛИТЕРАТУРА

1. Quel est le Nombre d'Utilisateurs Discord? (Chiffres Clés 2023). – Текст : электронный // L'Entrepreneur en vous. – URL: <https://lentrepreneurenvous.com/nombre-utilisateur-discord/> (дата обращения: 04.04.2024).

2. Discord statistics 2025 – how many people use Discord? – Текст : электронный // LEVVVEL. – URL: <https://levvvel.com/discord-statistics/> (дата обращения: 02.03.2024).
3. How Many People Use Discord? (2023-2024). – Текст : электронный // Bankmycell. – URL: <https://www.bankmycell.com/blog/number-of-discord-users/> (дата обращения: 18.04.2024).
4. Букина Л. М. Лингвистический и паралингвистический аспекты интернет-дискурса (на примере французских пользователей) – Текст : непосредственный // Вестник РУДН. Серия: Теория языка. Семиотика. Семантика. 2016. № 4. С. 125–134.
5. Яскевич, Д. М. Французский сетевой дискурс как социальная практика / Д. М. Яскевич, Л. П. Рыжова // Язык и действительность. Научные чтения на кафедре романских языков им. В. Г. Гака : Сборник статей по итогам VII международной конференции, Москва, 21–25 марта 2022 года. Том 7. – Москва: ООО "Издательство "Спутник+", 2022. – С. 521-524.
6. Distribution of Discord.com users worldwide as of February 2024, by age group – Текст : электронный // Discord Statistics – URL: https://translated.turbopages.org/proxy_u/en-ru.ru.f06bcb29-6866948b-5c5037bb-74722d776562/https/www.statista.com/statistics/1327674/discord-user-age-worldwide/ (дата обращения: 15.06.2024).
7. Cardon, D. Culture numérique / D. Cardon. – Paris, Les Presses de Sciences Po, 2019. – 432 p.

Секция 5. ЦИФРОВЫЕ ТЕХНОЛОГИИ В ПЕРЕВОДЕ И ОБУЧЕНИИ ПЕРЕВОДУ

УДК 81-26 347.78.034

Беклемешева Наталья Николаевна
кандидат филологических наук,
доцент Института иностранных
языков
Московский городской
педагогический университет
г. Москва, Россия

Natalia Beklemesheva
PhD in Philology,
Associate Professor at the Institute
of Foreign Languages
Moscow City University
Moscow, Russia
beclemesheva@mail.ru

РУССКИЕ ДЕЕПРИЧАСТИЯ И СПОСОБЫ ИХ ПЕРЕДАЧИ НА АНГЛИЙСКИЙ ЯЗЫК: ПРОФЕССИОНАЛЬНЫЙ ПЕРЕВОД VS МАШИННЫЙ (на материале художественной литературы)

В статье рассматриваются способы передачи русских деепричастий на английский язык на примере перевода рассказов А. П. Чехова. Рекомендуемые стратегии перевода деепричастий верифицируются в процессе сопоставления вариантов профессионального и машинного перевода. Анализ переводческих решений в работе основывается на выявлении функции деепричастия в тексте, оценка перевода учитывает сохранение идиостиля Чеховской прозы. Результаты исследования показали, что развертывание деепричастия в однородное сказуемое, а также сохранение при переводе полупредикативности с использованием английского причастия являются универсальными стратегиями перевода, прежде всего деепричастия в функции второстепенного сказуемого. Однако если деепричастие вносит в текст обстоятельственные характеристики, предикативность при переводе может опускаться, а функция деепричастия передаваться в английском языке с помощью наречий, именных групп и прилагательных. Машинный перевод, в отличие от профессионального, демонстрирует ограниченные возможности с точки зрения структурных преобразований и не способствует сохранению идиостиля художественного произведения.

Ключевые слова: профессиональный и машинный перевод; перевод русских деепричастий; перевод с русского языка на английский; перевод художественной литературы; вторичная предикативность.

RUSSIAN VERBAL ADVERBS IN FICTION: HUMAN TRANSLATION VS MACHINE TRANSLATION

The paper investigates the ways of rendering verbal adverbs in Russian fiction into English. The empirical corpus is based on A. Chekhov's stories and their English translation made by professional translator and Google translator. First, Russian verbal adverbs' functions in the text were analyzed, then translations were compared and assessed in terms of their pragmatic effect and the author's style. The findings confirm the most common translation strategies to render Russian verbal adverbs are: using participle I and predicate (finite verb form), which work perfectly if a converb functions as a secondary predicate in a sentence. Still, when an adverbial verb injects into fiction adverbial characteristics, it allows for significant translation

transformations, with a predicative component left out in favour of adverbs, adjectives and noun groups. Yet, machine translation potential for structural transformations appears to be rather limited; along with it, individual author's style might be lost on the foreign reader.

Key words: professional and machine translation; Russian adverbial verb translation; translation from Russian into English; literary translation; secondary predication.

Идея о том, что для каждого языка характерны свои предпочтения в употреблении структур-носителей предикативности, не нова [1; 2; 3]. Предикативность очевидно имеет свою лингвокультурную специфику, которая прежде всего проявляется в предпочтениях, которые отдельный язык проявляет в выборе структур-носителей предикативности разного «объема», т. е. структур-носителей полной (выражается личной формой глагола) и вторичной предикативности (полупредикативные, носители свернутой предикативности и в английском языке (далее АЯ) – скрытой предикативности). Вторичная предикативность, как известно, представляет собой такое описание денотативной ситуации (приписывание признака объекту), когда языковые признаки предикативности – модальность, время – выражены не полностью или опускаются, т. е. редуцируются (поэтому вторичная предикативность по сути является редуцированной предикативностью), ср.: **Он вошел и включил свет** (полная предикативность) vs **Войдя в комнату, он включил свет** (вторичная предикативность). Или: **He talked her into participating** (скрытая предикативность) vs **He talked with her and persuaded to / made her participate** (полная предикативность). В русском языке (далее РЯ), например, практически отсутствуют носители скрытой предикативности (*помог с докладом* vs *помог сделать/подготовить/проверить доклад*), в отличие от английского языка (*helped her into the car; let the dog out; decided against the dress*); зато присутствуют такие полупредикативные структуры, как деепричастия, которых нет в АЯ.

Степень редукции признаков предикативности определяется и жанровыми конвенциями. В РЯ, например, в текстах книжного стиля (официально-деловой, научный) существует ярко выраженная тенденция к употреблению структур-носителей свернутой предикативности (отглагольных существительных), в художественной литературе частотны полупредикативные структуры – причастные и деепричастные обороты. Эксперименты, проведенные исследователями, показывают, что при оценке текстов носителями языка как написанных на русском либо переведенных на русский язык с английского важную роль играет количество причастных и деепричастных оборотов в тексте. При отсутствии или наличии в тексте минимального количества причастных/деепричастных оборотов текст чаще воспринимается как переведенный с иностранного языка [4].

Иными словами, даже при наличии в языках сходных видов предикативных структур (причастные обороты, отглагольные существительные, инфинитивные структуры, сложные определения с внутренней предикацией) их перевод представляет определенную трудность в силу различной частотности данных единиц в паре РЯ и АЯ, обусловленных различными факто-

рами. Особую проблему с точки зрения перевода представляют структуры-носители предикативности, которые представлены только в одном языке, например, русское деепричастие, что составляет актуальность данной работы.

Специфика деепричастия как самостоятельной части речи и способы перевода деепричастий как носителей вторичной предикативности вызывает неизменный интерес исследователей [5; 6; 7; 8; 9], в том числе в свете активного использования систем машинного перевода в переводе в целом [10; 11]. Наряду с этим, некоторые вопросы до сих пор не получили детального описания и нуждаются в уточнениях, например: Существуют ли постоянные равнозначные переводческие соответствия между русскими деепричастиями и английскими структурами? Могут ли выявленные закономерности быть положены в основу машинного перевода? Возможно ли использовать машинный перевод для оптимизации переводческого процесса в отношении художественной литературы?

Цель исследования заключается в описании специфики передачи русского деепричастия на английский язык на примере художественной литературы и оценки потенциала машинного перевода в решении данной задачи.

Как хорошо известно, русское деепричастие представляет собой уникальную структуру, которая соединяет в себе признаки глагола и наречия. Внося в предложение информацию о втором признаке субъекта матричного предложения, т. е. дополнительную предикативность, деепричастие выполняет в предложении функцию второстепенного сказуемого. Важной характеристикой данной структуры является указание на фазу действия: дополнительное действие может происходить одновременно с действием сказуемого, предшествовать ему или осуществляться после него. Несмотря на термин «второстепенное» сказуемое, деепричастие не всегда можно отнести к статусу «периферийной» пропозиции, ср.: *Он сидел, оглядываясь по сторонам* vs *Он сидел и оглядывался по сторонам*.

Вторая важная функция, которую выполняет деепричастие – это указание на различные обстоятельственные характеристики, сопровождающие основное действие. Исследователи выделяют следующие виды обстоятельств, ср.:

- образа действия: *повышая голос; улыбаясь* (описание языка тела) vs *нервничая* (указание на эмоциональное состояние);
- причины: *она вскрикнула, испугавшись*;
- цели: *остановился, отдыхая*;
- условия: *выполнив задания в срок, вы улучшите свои баллы*;
- сравнения: *она вела себя странно, словно забыв обо всем*;
- уступки: *хоть и прихрамывая, он справился быстрее*;
- времени: *пообедав, они продолжили путь* [7; 8; 9].

Способы перевода деепричастия, описываемые на сегодняшний день в пособиях по переводу, сводятся в основном к двум рекомендациям: развертывание деепричастия в полную предикативность – второе сказуемое, выраженное личной формой глагола, или использование английских причастий и герундиальных оборотов, при этом виды причастий зависят от грамма-

тической связи деепричастия со сказуемым, ср.: *having done (сделав) vs doing (делая)*. Это, предположительно, самые частотные и эффективные стратегии перевода деепричастий и деепричастных оборотов, однако задача перевода в данном случае осложняется нюансами логических связей между деепричастием и сказуемым. Кроме того, при переводе художественной литературы необходимо учитывать прагматический потенциал деепричастия.

Прежде всего, способ передачи русского деепричастия / деепричастного оборота (далее употребляется недифференцировано) на АЯ зависит от его функции в предложении, и акцент может быть сделан либо на передаче предикативности, либо на обстоятельственных характеристиках.

Перевод деепричастий **в функции второго сказуемого** основан на передаче предикативности, поскольку деепричастие указывает действие, равное по значимости действию глагола-сказуемого. Именно поэтому адекватным переводом в данном случае является развертывание деепричастия в однородное сказуемое, ср.: *Сказав это, он вышел из комнаты* – ‘He **said** it and left’. Важную роль при выборе переводческого эквивалента играют временные отношения описываемых в предложении ситуаций: если деепричастие указывает на действие, предшествующее действию сказуемого, оно может передаваться на английский язык с помощью: 1) перфектной формы английского причастия: ‘**Having said it**, he left; 2) герундием с предлогом (*after, before doing*): *Поговорив с адвокатом, он убедился, что может выиграть дело* – ‘**After talking** with the lawyer he understood he had good chances to win’; 3) а также полной предикативной структурой: *Выйдя на широкий проспект* – ‘As we **came on** to the broad avenue’. 4) Помимо этого, в тех случаях, когда второе действие, выраженное деепричастием, указывает на объект (*выполнив работу; оставив сообщение*), эффективной стратегией может быть использование Абсолютной конструкции (*with smth done*). В таком случае при переводе меняется перспектива описания ситуации и дополнительное действие описывается с помощью акцента на объекте, который становится субъектом независимого причастного оборота, ср.: *Сделав свое дело, они давно ушли* – ‘(With) **their work / mission completed**, they left’. *Оставив сообщение, он вышел на улицу* – ‘(With) **the message left**, he was out on the pavement’.

Одновременное и последующее действие, выраженное деепричастием, традиционно передается на АЯ с помощью причастия I / герундия (возможно использование союзов *while/when*) или путем развертывания в полную предикативность, ср.: *Сказал он, краснея/покраснев* – ‘He said, **blushing / and turned red**’.

При этом, если последующее действие является неожиданным (и часто неприятным для субъекта), при переводе русское деепричастие можно передать английским инфинитивом последующего действия, ср.: *Он открыл дверь, обнаружив за ней длинный темный коридор* – ‘He opened the door **to find** a long dark corridor’. В данной структуре используется деепричастие, образованное от глаголов с семантикой чувственного восприятия (*увидев / услышав / почувствовав запах* – *to see, hear, feel, taste, smell*); обнаружения (*открыв, столкнувшись, найдя* – *to find, discover, come across, bump into*);

а также от пассивной формы глагола – действие, которое трудно предугадать, соответственно, оно тоже может являться неожиданным для субъекта, ср.: *he came in to be received with a bear hug*.

В том случае, если деепричастие выполняет обстоятельственную функцию, стратегия перевода во многом зависит от типа обстоятельства.

При передаче на АЯ деепричастия в функции обстоятельства причины / цели / условия сохраняется курс на передачу предикативности: при переводе используется английское причастие или полная предикативность (второе сказуемое), поскольку деепричастие указывает на совершение действия 1, послужившего поводом для последующего действия 2, ср. обстоятельство причины: *он отвернулся, не желая видеть* – ‘he turned away as he **did not want to see it / unwilling to see it**’. Или: *вынашивая план мести, он забыл о главном* – ‘**Plotting** revenge, he forgot something important’. В функции цели деепричастие чаще всего переводится с помощью инфинитивного оборота цели: *улаживая конфликт, он предложил встретиться и обсудить разногласия* – ‘**to handle the conflict** he suggested first discussing disagreements in person’. Ср. также обстоятельство условия: *Выучив все как следует, ты сможешь улучшить свой балл* – ‘**If you learn** everything properly...’

Стратегия сохранения исходной предикативности при переводе (с использованием английского причастия/герундия) или развертывания деепричастия в полную предикативность (английское однородное сказуемое) представляется универсальной. Тем не менее, при передаче на другой язык обстоятельственной характеристики образа действия у переводчика появляется бóльшая свобода маневра, поскольку главный акцент делается на состоянии и поведении субъекта, что требует передачи прагматического аспекта значения деепричастия. В том случае, если описывается язык тела (*Он шептался о чем-то с коллегами, нервно поглядывая на дверь*), можно использовать причастие 1 (‘He was whispering about something with his colleagues, **looking nervously** at the door’); Абсолютную конструкцию (*Он сидел, сложив руки на коленях* – ‘He was sitting, hands on his knees’); или же метонимически преобразованный глагол (по модели обстоятельственной метонимии, когда действие описывается через обстоятельственные характеристики), ср.: *сказал, запинаясь* – ‘**mumbled/stuttered**’.

Если главной задачей автора является описание с помощью деепричастия состояния героя, предикативность (деепричастие) может опускаться в пользу наречия, именной группы или прилагательного, ср.: *Сказала мать, сделав строгое лицо* – ‘said mother **sternly / in a stern manner / her face stern**’ (Абсолютная конструкция).

Деепричастия в функции обстоятельства времени (и иногда обстоятельства места) используются для указания на время / место совершения главного действия, в связи с чем, помимо личной формы глагола и причастия, при переводе можно использовать только именную группу, ср.: *вернувшись домой, он сделал звонок* – ‘**back home** he made a call’; *пообедав, мы продолжили работу* – ‘**after dinner / dinner over**, we sat down to work again’.

Данная стратегия может использоваться и при переводе деепричастия в функции сравнения, ср.: *напоминая два растопыренных пальца* – ‘like the spread fingers of a hand’.

Стратегии, предложенные в работе, основываются прежде всего на предпочтениях языков разных типов. Однако при переводе художественного произведения важными представляются и другие критерии, например, сохранение идиостиля автора, что делает задачу сохранения исходной структуры при переводе важнее, чем совершение трансформаций, обусловленных разницей языков [4].

Для того, чтобы выявить наиболее частотные стратегии передачи русских деепричастий на АЯ и причины, стоящие за выбором переводчика, был выполнен анализ художественного текста и его перевода на английский язык. Источником эмпирического материала послужили рассказы А. П. Чехова («Анна на шее», «Гриша», «Дачники», «Дорогие уроки», «Шуточка», «Произведение искусства») и их перевод на английский язык (пер. Е. В. Полищук, 2015 г.) [12].

С помощью ручной выборки из оригинального текста, объем которого составил в целом 14 718 слов, было отобрано 108 контекстов с деепричастиями. Далее все отобранные контексты были разделены на группы в соответствии с видом деепричастия (совершенный vs несовершенный) и функциями (второстепенное сказуемое, обстоятельство образа действия, причины, цели, сравнения, условия).

Анализ корпуса примеров показал, что деепричастия несовершенного типа используются гораздо чаще, чем совершенного: 81,5 % vs 18,5 %. Самой распространенной функцией в корпусе примеров является функция второстепенного сказуемого (63 %), следующей по частотности – обстоятельство образа действия (33 %); остальные обстоятельственные характеристики носят в корпусе примеров единичный характер: обстоятельство сравнения (2)¹, обстоятельство причины (1), обстоятельство цели (1).

На следующем этапе были выявлены стратегии передачи деепричастий и деепричастных оборотов на английский язык в профессиональном переводе.

Анализ показал, что деепричастие совершенного вида в функции второстепенного сказуемого (*усевшись*) передается в профессиональном переводе с помощью: 1) причастия I – *sitting* (6); перфектного причастия – *having sat* (4); *after* + причастие I – *after taking his place* (4) и, наконец, личной формы глагола (второе сказуемое) – *he sat down and ...* (6).

Деепричастие несовершенного типа (*говоря*) передается на английский в подавляющем большинстве случаев с помощью причастия I – *saying* (68 %), и лишь в шести (6) случаях деепричастие при переводе разворачивается в полную предикативность – личную форму глагола (*said*).

¹ в скобках указано количество примеров

Среди деепричастий в обстоятельственной функции в рассказах А. П. Чехова были выявлены: деепричастие в функции условия (1) – при переводе разворачивается в полную предикативность; сравнения (1) – разворачивается в полную предикативность; деепричастие образа действия (22). В последнем случае стратегии перевода более разнообразны. Анализ переводов показал, что деепричастия и деепричастные обороты в этой функции передаются на английский язык:

1) причастием I, ср.: *задыхаясь и запинаясь, стал объясняться в любви* (Дорогие уроки) – ‘**panting and stammering**, he began to declare his love’ (8);

2) наречием (при этом происходит опущение предикативности), ср.: *начал он, задыхаясь* (Произведение искусства) – ‘he began **breathlessly**’ (3);

3) личной формой глагола (развертывание деепричастия в полную предикативность), ср.: *Пожирая глазами свою учительницу* (Дорогие уроки) – ‘He **greedily ate** his delicate teacher **with eyes**’; в том числе и с помощью метонимически преобразованного глагола (по модели обстоятельственной метонимии), ср.: *сказал, задыхаясь* – ‘he **gasped**’ (дышать с трудом, затаить дыхание, охнуть) (2);

4) прилагательным, ср.: *Она улыбалась, волнуясь при мысли* (Анна на шее) – ‘She returned the smile, **nervous** at the idea’ (3);

5) Абсолютной конструкцией, ср.: *ходили за ним, опустив свои острые морды* (Анна на шее) – ‘were walking **with their heads lowered**’ (1).

6) Наконец, в нескольких случаях переводчики опускают деепричастие при переводе, ср.: *пробормотал он, смущаясь еще больше* (Дорогие уроки) – ‘he muttered to himself’ (5).

Выбор профессиональным переводчиком стратегий передачи деепричастий на английский язык демонстрирует явные предпочтения, во-первых, это сохранение полупредикативности в виде причастия I. По всей видимости, это решение обусловлено стремлением максимально близко к оригиналу передать идиостиль А. П. Чехова: длину предложений и их ритм. Остальные переводческие решения носят единичный характер. Особый интерес представляют случаи опущения предикативности (глагольных форм) – как правило, это происходит в случае описания эмоционального состояния героев, которое в английском языке осуществляется с помощью прилагательных и наречий, а также именных групп. Важно также выяснить, в каких случаях переводчик может отказываться от передачи деепричастий в художественном тексте. Всего в эмпирическом корпусе было выявлено пять случаев опущения исходного деепричастия / деепричастного оборота, ср.: 1) *Укладываясь на ночлег, шумно кричат грачи* – ‘We can hear some young rooks cawing in the background’ (букв. *Мы слышим грачей, каркающих вдалеке*); 2) *...пробормотал он, смущаясь еще больше...* (Дорогие уроки) – ‘he muttered to himself’; 3) *Она взяла; но, заказывая бальное платье, ни с кем не советовалась* (Анна на шее) – ‘She took it and didn’t consult with anybody about her gown’; 4) *Переходя с рук на руки, угорая от музыки и шума, мешая русский язык с французским, картавя, смеясь и не думая ни о муже, ни о ком*

и ни о чем (*Анна на шее*) – ‘Going from one partner to another, using Russian and French, laughing and not thinking at all about her husband or about anything else’;
5) ...мужчины вытянулись как-то странно, **опустив руки**... (*Анна на шее*) – ‘and the men came to attention’.

Опущение деепричастного оборота в примере 1 можно объяснить невысокой коммуникативной значимостью данного компонента предложения. Сказуемое матричного предложения описывает крики скворцов, указание на процесс их укладывания на ночлег опускается, поскольку из контекста понятно, что это вечер, солнце садится. Схожие аргументы в пользу опущения деепричастия можно предложить и в примере 3. Здесь деепричастие также выступает в роли второстепенного сказуемого, которое указывает на действие, очевидное из контекста. В примерах 2 и 4 деепричастие используется в обстоятельственной функции. В примере 2 эмоциональное состояние героя передается с помощью уточнения характера звука и речевого действия в целом: *muttered to himself* – букв. *пробормотал себе*, т. е. *под нос*, что естественно в состоянии крайнего смущения. Решение переводчика в примере 4 скорее направлено на снижение экспрессивности исходного фрагмента: опускаются деепричастия, описывающие эйфорию героини и особенности ее речи. В примере 5 деепричастный оборот переосмысливается переводчиком как часть общей ситуации – приветствие вышестоящего по статусу чиновника, и подвергается метонимическому преобразованию: *вытянулись, опустив руки* => встали по стойке «смирно».

На следующем этапе отобранные фрагменты с деепричастными оборотами были переведены с помощью платформы машинного перевода Гугл переводчик.

Частотность случаев передачи деепричастия несовершенного вида с помощью причастия I практически совпадает с решениями профессионального переводчика (65 %).

Деепричастия совершенного вида также передаются с помощью perfectной формы причастия (8); личной формы глагола (5); структуры *after + doing* (2); причастия I (4); в одном случае используется Абсолютная конструкция, что в принципе несколько отличается от профессионального перевода, при котором очевидное предпочтение даже при переводе деепричастия совершенного вида отдается причастию I.

Основные расхождения относятся к передаче обстоятельственной функции деепричастия. В машинном переводе было выявлено только два случая опущения предикативности: *Доктор! – начал он, задыхаясь. – Представьте мою радость!* (Произведение искусства) – “Doctor!” he began, **out of breath**. “Imagine my joy!” и *Он ворочается с боку на бок, болтает и в конце концов, не вынося своего возбуждения, начинает плакать* (Гриша). ‘He tosses and turns, chatters and finally, **unable** to bear his excitement, begins to cry’. В обоих случаях система опускает причастие I, образованное от глагола *to be (being)*, в предикатах *to be out of breath* и *to be unable*. Однако более серьезных трансформаций, связанных с метонимическими переносами, в машинном

переводе найти не удалось. В основном предлагаются дословные варианты перевода с минимальными трансформациями. Ср. профессиональный перевод: ...*продолжал он, **впадая в торжественный тон*** – ‘he continued **solemnly**’ (букв. *продолжал торжественно*) vs машинный: ‘he continued, suddenly **falling into a solemn tone**’. Или: ...и мужчины вытянулись как-то странно, *опустив руки...* – ‘the men **came to attention**’ (профессиональный перевод) vs ‘the men stretched out in a strange way, **with their hands hanging down**’ (Гугл переводчик).

В машинном переводе не было отмечено ни одного случая опущения деепричастного оборота. Рассмотрим следующий фрагмент рассказа А. П. Чехова «Дорогие уроки»: *Но почему-то в антрактах у него билось сердце, он, **сам того не замечая**, как мальчик бегал по фойе и по коридорам, нетерпеливо отыскивая кого-то*. Профессиональный переводчик не передает значение деепричастного оборота, очевидно считая его избыточным: ‘But for some reason during the intermission his heart was beating heavily and he rushed to the foyer like a young bow impatiently looking for someone’, а Гугл переводчик сохраняет: ‘... and **without noticing it**, he ran like a boy through the foyer and the corridors, impatiently looking for someone’.

Ср. также переводы еще одного фрагмента: *Она танцевала страстно, с увлечением и вальс, и польку, и кадрили, переходя с рук на руки, **угорая от музыки и шума**, мешая русский язык с французским, **картавя**, смеясь и не думая ни о муже, ни о ком и ни о чём* (Анна на шее) – ‘She danced the waltz, polka and the quadrille, and **with delight**, going from one partner to another, using Russian and French, laughing and not thinking at all about her husband or anything else’ (пер. Е. В. Полищук) vs ‘She danced passionately, with enthusiasm, the waltz, the polka, and the quadrille, moving from hand to hand, **overcome with joy from the music and noise**, mixing Russian with French, **lispering**, laughing, and not thinking about her husband, or anyone, or anything’ (Гугл переводчик). В профессиональном переводе эмоциональное состояние героини описывается с помощью именной группы (*with delight*), деепричастие как носитель предикативности опускается; во втором случае *картавя* опускается совсем, поскольку ранее упоминается, что героиня говорила на французском. В машинном переводе предикативность сохраняется в обоих случаях: *lispering* и *overcome* (букв. *охваченная музыкой и шумом*), однако это значительно утяжеляет английский перевод, ср.: оригинальное предложение составляет 2,5 строчки, а в переводе Гугл – более трех.

Еще одно очевидное различие в исследуемых переводах относится к передаче отрицательных деепричастий. Если машинный перевод использует только одну переводческую модель – *without+doing* (герундий), то в профессиональном переводе стратегии гораздо более разнообразны. Прежде всего, наиболее частотна модель *not doing* (причастие I) (50 %), ср.: *не откладывая дело в долгий ящик* – ‘**not putting** the deal aside’, в отличие от модели *without doing* (1). Переводчик также обращается к антонимичному переводу (*не*

отрывая глаз с Ани – **looking at Anya**), а также к развертыванию в полную предикативность (он машет рукой и, **не ответив** на вопрос, заводит речь о погоде – ‘... **doesn’t answer** the question and changes the subject to the weather’. Выбор профессионального переводчика в пользу менее распространенной структуры (*not doing*) вместо *without doing* можно также отнести к стремлению сохранить длину предложения при переводе и в целом стиль чеховской прозы.

Таким образом, исследование показало, что выбор английских соответствий при переводе деепричастия, несомненно, зависит прежде всего от его функции в русском предложении. Тем не менее, смещение акцентов с передачи предикативности на передачу обстоятельственных характеристик пока под силу только профессиональным переводчикам. Машинный перевод может помочь в процессе подбора эквивалентов на уровне лексики, однако далее требует обязательного осмысления авторского замысла профессиональным переводчиком с последующими метонимическими и структурными преобразованиями.

ЛИТЕРАТУРА

1. Грамматические аспекты перевода / О. А. Сулейманова, Н. Н. Беклемешева, К. С. Карданова-Бирюкова, Н. В. Лягушкина. 2012. 240 с.
2. Сулейманова О. А., Карданова-Бирюкова К. С., Контрастивный анализ русских и английских «периферийных» синтаксических структур в переводческой перспективе // Русистика и компаративистика: Сборник научных трудов по филологии. 2021. № 15. С. 297–315.
3. Беклемешева Н. Н., Английские структуры-носители вторичной предикативности с точки зрения перевода // Исследования языка и современное гуманитарное знание. 2022. Т.4, № 2. С. 91–100.
4. Иваненко К. В. Преобразования предикативных структур при переводе в паре «русский – английский» (на материале научно-популярных текстов) // Вестник МГЛУ. Гуманитарные науки. 2024. Вып. 7 (888). С. 51–58.
5. Абдулхакова Л. Р. О грамматическом статусе деепричастия // Вестник Нижегородского университета им. Н.И. Лобачевского. 2010. № 4 (2). С. 415–418.
6. Алексеева Я. В. Синтаксические функции деепричастных форм в разножанровых текстах второй половины XVIII века // Ученые записки Казанского ГУЮ Гуманитарные науки. 2010. Т. 19, Кн. 2. С. 173–181.
7. Кудрявцева Т. Ю. Второе сказуемое, репрезентированное деепричастием, в системе предикативных средств // Филологические науки. Вопросы теории и практики. 2013. № 8–1. С. 96–99.
8. Ди Л. Прагматические аспекты перевода русского деепричастия на китайский язык // Мир русского слова. 2016. № 2. С. 19–25.
9. Ермолович Д. И. Русско-английский перевод. Москва : Аудитория, 2022. 640 с.

10. Suleimanova O. A. The Learning and Educational Potential of Digital Tools in Humanities and Social Science // Society. Integration. Education: Proceedings of the International Scientific Conference, Rēzekne, 22–23 мая 2020 года. Vol. 4. 2020. P. 657–669.
11. Петрова И. М., Иванова А. М., Никитина В. В. Современные цифровые технологии в лингвистических исследованиях. Московский городской педагогический университет, Институт иностранных языков. Москва : Языки Народов Мира, 2022. 259 с.
12. Полищук Е. В., Суровцева Е. В. Чехов / Chekhov. Рассказы А. П. Чехова по-русски и по-английски с комментариями. Москва : Р.Валент, 2015. 94 с.

УДК 378.4

Берендяев Максим Викторович

доцент базовой кафедры технологий
и автоматизации перевода в бюро
переводов «АКМ-Вест»

Северный (Арктический) федеральный
университет имени М. В. Ломоносова
г. Архангельск, Россия

Maxim Berendyaev

Associate Professor of the Department
of Translation Technology
and Practice at AKM-West

Northern (Arctic) Federal University
Arkhangelsk, Russia
m.berendyaev@narfu.ru

Коканова Елена Сергеевна

кандидат филологических наук,
заведующий базовой кафедрой
технологий и автоматизации перевода
в бюро переводов «АКМ-Вест»

Северный (Арктический) федеральный
университет имени М. В. Ломоносова
г. Архангельск, Россия

Elena Kokanova

PhD in Philology,
Head of the Department of Translation
Technology and Practice at AKM-West
Northern (Arctic) Federal University

Arkhangelsk, Russia
e.s.kokanova@narfu.ru

ПЕРСПЕКТИВЫ ДАЛЬНЕЙШЕГО РАЗВИТИЯ КЛАССИЧЕСКИХ САТ-ИНСТРУМЕНТОВ ДЛЯ ПЕРЕВОДА ПО МНЕНИЮ ПРОФЕССИОНАЛЬНОГО СООБЩЕСТВА

В статье делается обзор современных систем автоматизированного перевода в контексте их эволюции и влияния на профессиональную переводческую деятельность. Исследование фокусируется на выявлении универсальных закономерностей, характерных как для зарубежных, так и для отечественных САТ-инструментов, избегая узкоспециализированной оценки конкретных программных решений. Актуальность работы обусловлена необходимостью переосмысления функциональных возможностей САТ-систем в условиях растущих требований к качеству, гибкости и безопасности современных технологий перевода и локализации. В статье выделены ключевые аспекты, требующие улучшения: автоматизация и интеграция, удобство и интуитивность, качество данных, надежность и безопасность, а также гибкость и открытость. Особое внимание уделяется потенциалу нейросетевых технологий и облачных вычислений для преодоления существующих ограничений. Перспективным направлением исследований признается разработка унифицированной методологии оценки эффективности САТ-инструментов. Проведенный анализ предлагает пути совершенствования САТ-систем в соответствии с современными тенденциями цифровой трансформации лингвистических услуг.

Ключевые слова: автоматизированный перевод; профессиональная переводческая деятельность; САТ-инструменты; нейросетевые технологии; оценка эффективности САТ-инструментов; цифровая трансформация лингвистических услуг.

FUTURE DEVELOPMENT PROSPECTS OF CLASSICAL CAT TOOLS: THE PROFESSIONAL COMMUNITY'S PERSPECTIVE

The paper provides an overview of modern computer-aided translation systems in the context of their evolution and impact on professional translation practices. The study focuses on identifying universal patterns common to both foreign and Russian CAT tools, avoiding

a narrowly specialized assessment of specific software solutions. The relevance of the work stems from the need to reconsider the functional capabilities of CAT systems amid growing demands for quality, flexibility, and security in modern translation and localization technologies. The paper highlights key areas requiring improvement: automation and integration, usability and intuitiveness, data quality, reliability and security, as well as flexibility and openness. Special attention is paid to the potential of neural network technologies and cloud computing in overcoming existing limitations. The development of the unified methodology for assessing CAT tool efficiency is recognized as a promising research direction. The conducted analysis suggests ways to enhance CAT systems in line with current trends in the digital transformation of linguistic services.

Key words: computer-aided translation; professional translation practices; CAT tools; AI-based language technologies; performance assessment of CAT tools; digital transformation of linguistic services.

Настоящее исследование направлено на анализ современных систем автоматизированного перевода (Computer-Aided Translation, Computer-Assisted Translation, CAT), рассматриваемых в контексте их эволюции и влияния на профессиональную переводческую деятельность. В отличие от узкоспециализированных оценок конкретных программных решений, в статье делается фокус на выявлении универсальных закономерностей, присущих как зарубежным, так и отечественным САТ-инструментам.

За последние два десятилетия технологии автоматизированного перевода существенно трансформировали лингвистическую практику, внедрив новые стандарты обработки текста и оптимизации переводческих процессов. Однако текущий этап развития САТ-систем характеризуется замедлением инновационной динамики, что свидетельствует о достижении определенного технологического плато. Несмотря на очевидные преимущества, включая ускорение рутинных операций и стандартизацию рабочих процессов, профессиональные лингвисты продолжают сталкиваться с рядом системных ограничений. Эти барьеры препятствуют полной реализации потенциала автоматизации в таких сферах, как переводческая деятельность, лингвистический анализ и локализация.

Таким образом, актуальность настоящего исследования обусловлена необходимостью переосмысления функциональных возможностей САТ-систем в условиях развития технологий искусственного интеллекта (ИИ), растущих требований к качеству, гибкости и безопасности лингвистических технологий. Рассмотрим ключевые аспекты, требующие, на наш взгляд, улучшения.

1. Автоматизация и интеграция

Ключевой проблемой современных САТ-инструментов остается недостаточный уровень автоматизации рутинных операций. Повышение рыночной конкурентоспособности переводческих инструментов требует реализации следующих функциональных возможностей: интеллектуального подбора переводческих кадров на основе специализации и рейтинговых показателей;

комплексной интеграции с фриланс-платформами, платежными системами, CRM-решениями; автоматизированного подключения к лексикографическим ресурсам, терминологическим базам, системам автоматизации перевода. Существующая необходимость постоянного участия проектных менеджеров в данных процессах существенно снижает операционную эффективность. Методологическое решение предполагает создание централизованных платформенных решений, включающих усовершенствованную систему управления переводческой памятью (ТМ) с поддержкой модернизированного ТМХ-формата, обязательной верификацией метаданных, атрибуцией авторства; терминологические базы данных, основанные на усовершенствованном стандарте TBX, автоматизированных механизмах синхронизации.

Анализ существующих решений выявляет существенные недостатки в процессах импорта / экспорта данных, механизмах межсистемной синхронизации, что приводит к необходимости разработки кастомных скриптов (Python и др.), ручного редактирования в табличных процессорах.

С экономической точки зрения особую значимость приобретает автоматизация конвертации единиц измерения в переводческой отрасли [1]: прогнозирование объема перевода, расчет стоимостных показателей, адаптивного сокращения текстов (лингвистические метрики, математические модели оптимизации).

Создание интегрированных платформенных решений позволит оптимизировать ресурсозатраты, повысить точность планирования, улучшить управление переводческими и локализационными проектами.

2. Удобство и интуитивность

Когнитивная перегрузка пользователей современных САТ-систем представляет собой значительную проблему в области компьютерной лингвистики. Проведенный анализ выявил следующие системные недостатки интерфейсных решений: избыточная функциональная насыщенность, создающая барьеры для начинающих пользователей, отсутствие дифференциации интерфейса по уровням профессиональной подготовки, дефицит встроенных обучающих систем и контекстно-зависимых подсказок. Особого внимания требует вопрос обработки лингвистически сложных языковых систем. Существует необходимость разработки специализированных алгоритмов для неалфавитных языков, языков с нетипичной синтаксической структурой, создания адаптивных моделей сегментации текста, усовершенствования методов расчета объема перевода.

Актуальность данных направлений обусловлена постепенной автоматизацией перевода между индоевропейскими языками, сохранением актуальности гибридной модели «человек-машина» для малоресурсных языков, языков с уникальной морфосинтаксической структурой.

Кроме того, развитие классических САТ-инструментов позволяет систематизировать требования к интерфейсам САТ-систем, разработке методоло-

гии обработки сложных языковых систем, созданию основы для дальнейших исследований в области компьютерной лингвистики, автоматического перевода, человеко-машинного взаимодействия [2; 3].

3. Качество данных

Современные САТ-системы требуют перехода от пассивного хранения переводческих данных к интеллектуальной обработке лингвистического контекста. Ключевые направления развития включают контекстно-ориентированный анализ, обеспечивающий автоматическое выявление терминологических несоответствий, стилистический контроль, тематическую адаптацию (например, юридическую, медицинскую, техническую), оптимизацию автоматического перевода через анализ качественных параллельных корпусов.

Как уже отмечалось нами ранее, существуют определенные требования к разработчикам систем автоматизации перевода: при формировании текста не учитывается весь проект и (или) весь текст целиком, наблюдается сегментация и фрагментация на не связанные между собой сегменты [4, с. 9]. При работе с САТ-инструментами следует обращать внимание на необходимость преодоления архаичной сегментации текста: существующая практика фрагментарной обработки противоречит принципам целостного перевода, современные вычислительные ресурсы позволяют реализовать контекстно-зависимые модели. Для преодоления ограничений традиционной сегментации, совершенствования человеко-машинного взаимодействия необходимо формировать расширенные промпты для ИИ-систем, содержащие техническое задание проекта, стратегию перевода (на основе предпереводческого анализа), полный исходный текст, историю уже выполненных переводов в рамках того же текста или проекта (например, предыдущих фрагментов при сегментной работе), реализацию контекстно-зависимых моделей перевода. Такой подход соответствует современным тенденциям цифровизации лингвистических услуг и открывает новые возможности для автоматизации переводческой деятельности [5].

4. Надежность, производительность и безопасность

К системным проблемам в работе современных САТ-систем, существенно ограничивающих их операционную эффективность, можно отнести замедление обработки объемных файловых массивов, нестабильность синхронизации в облачных средах, потенциальные риски компрометации данных (особенно в корпоративном сегменте), а также потенциальные конфликты при одновременной работе в локальном клиентском программном обеспечении (так называемых «десктопных редакторах», «тонких клиентах») и браузерных решениях с единой серверной или облачной инфраструктурой. Для минимизации указанных рисков требуется внедрение многоуровневой системы защиты, включающей: детализированное логирование всех модификаций, сквозное шифрование данных на всех этапах обработки, систему резервного копирования с сохранением полной истории изменений, функциональные требования к архитектуре систем.

Реализация предложенной архитектуры позволит существенно повысить надежность САТ-систем и их соответствие современным требованиям информационной безопасности в профессиональной лингвистической деятельности.

5. Гибкость и открытость

Современные САТ-инструменты демонстрируют выраженную экосистемную замкнутость, что создает существенные ограничения для масштабируемости и адаптивности. В качестве методологического решения предлагается внедрение модульной архитектуры, предусматривающей гибкую конфигурацию функциональных компонентов, открытые API-интерфейсы для интеграции с системами управления терминологией (Terminology Management), инструментами контроля качества (QA Tools), поддержку пользовательских скриптов для автоматизации специализированных задач, многофакторной рейтинговой системой оценки качества, индекс глубины правки, коэффициент исправления ошибок [2].

Подчеркнем, что реализация предложенных решений позволит преодолеть существующие ограничения замкнутых экосистем и вывести автоматизацию переводческой деятельности на качественно новый уровень, соответствующий современным тенденциям цифровой трансформации лингвистических услуг.

Таким образом, проведенное исследование позволяет выявить принципы взаимодействия между переводчиком и «машиной». Это вносит определенный вклад как в современное переводоведение, так и в практику профессиональной переводческой деятельности. Идеальная САТ-система, согласно выявленным критериям, должна сочетать высокую степень автоматизации рутинных операций, интуитивно понятный интерфейс, минимизирующий время адаптации пользователя системы, контекстно-ориентированные алгоритмы обработки текста, надежную систему защиты лингвистических данных, гибкие механизмы кастомизации под индивидуальные ‘workflow’. Хотя современные решения еще не достигли оптимального сочетания всех указанных характеристик, наблюдаемая динамика развития технологий ИИ и облачных вычислений свидетельствует о постепенном приближении к этой цели. Особый интерес для дальнейших исследований представляет изучение потенциала нейросетевых технологий в преодолении существующих ограничений САТ-систем. Перспективным направлением представляется разработка унифицированной методологии оценки эффективности САТ-инструментов, которая позволила бы объективно сравнивать различные системы и прогнозировать тенденции развития данного сегмента переводческих технологий.

ЛИТЕРАТУРА

1. Стандарт АПК по единицам измерения в переводческой отрасли. URL: <https://atcru.org/upload/iblock/07e/yrt1meetq414ldamcbu28wtiup98tpqe.pdf> (дата обращения: 20.06.2025).

2. Siu S. Ch. Revolutionizing Translation with AI: Unravelling Neural Machine Translation and Generative Pre-Trained Large Language Models [Electronic resource]. 2023. DOI: <http://dx.doi.org/10.2139/ssrn.4499768>.
3. Zade N. Automatic text summarization and translation using NLP with its functioning in AI: a comprehensive review [Electronic resource]. 2024. DOI: <https://www.doi.org/10.56726/IRJMETS47585>.
4. Берендяев М. В., Светова С. Ю., Коканова Е. С. Автоматическая генерация текстов в профессиональном переводе // DIDACTICA TRANSLATORICA. 2024. №2. С. 5–10.
5. Стандарт ассоциации переводческих компаний по применению средств и постобработке результатов автоматизированной и автоматической генерации текста (память переводов, машинный перевод, большие речевые модели) в профессиональном переводе (стандарт АПК по АГТ) / Non-human text generation for augmented professional translation and interpretation services (TM, MT, LLM). Standard of the Russian Association of Translation Companies. Проект 2.1 от 24.04.2025 // PEMT. URL: https://pemt.ru/wp-content/uploads/2025/04/standard_apk_agt2.pdf (дата обращения: 29.06.2025).

УДК 81'322.4

Волкова Ирина Дмитриевна

кандидат филологических наук,
доцент кафедры теории и практики
перевода

Волгоградский государственный
университет
г. Волгоград, Россия

Irina Volkova

PhD in Philology,
Associate Professor of the Translation
Studies Department
Volgograd State University
Volgograd, Russia
romadina-irina@mail.ru

Новожилова Анна Алексеевна

кандидат филологических наук,
доцент кафедры теории и практики
перевода

Волгоградский государственный
университет
г. Волгоград, Россия

Anna Novozhilova

PhD in Philology,
Associate Professor of the Translation
Studies Department
Volgograd State University
Volgograd, Russia
novozilova@yandex.ru

Митягина Вера Александровна

доктор филологических наук,
профессор кафедры теории и практики
перевода

Волгоградский государственный
университет
г. Волгоград, Россия

Vera Mityagina

PhD in Philology,
Professor of the Translation Studies
Department
Volgograd State University
Volgograd, Russia
mityagina@mail.ru

Никитенко Татьяна Васильевна

кандидат филологических наук,
заведующий кафедрой теории
и практики перевода

Белорусский государственный
университет
г. Минск, Беларусь

Tatyana Nikitenko

PhD in Philology,
Head of the Translation Studies
Department
Belarusian State University
Minsk, Belarus
tatnikita@gmail.com

**СПЕЦИФИКА ПРЕДРЕДАКТИРОВАНИЯ ПРЕЗЕНТАЦИОННЫХ
ТЕКСТОВ ДЛЯ МАШИННОГО ПЕРЕВОДА
(на материале английского и немецкого языков)**

Статья посвящена исследованию особенностей процесса предредактирования презентационных текстов, размещаемых на сайтах различных организаций, для последующего перевода в системах МП. Актуальность данного исследования обусловлена частотностью практики использования сервисов МП для перевода вербального контента сайтов, как разработчиками сайтов, так и пользователями. Исследование проведено на материале контента одноязычных сайтов медицинских и культурно-досуговых учреждений Российской Федерации. С целью выявления технологии предредактирования использована система машинного перевода Google Translate, и проанализирован перевод

презентационных текстов, выполненный данной системой на английский и немецкий языки. В результате проведенного анализа были выявлены ошибки, допускаемые системой Google Translate при работе с языковыми парами русский-английский, русский-немецкий, и предложены техники редактирования исходного текста с учетом данных ошибок.

Ключевые слова: предредактирование; презентационный текст; машинный перевод; многоязычный сайт; постредактирование; исходный текст.

PRE-EDITING OF PRESENTATIONAL TEXTS FOR MACHINE TRANSLATION (Based on the English and German Languages)

The article explores the pre-editing of presentational texts published on the websites of various organizations, intended for subsequent translation using machine translation (MT) systems. The relevance of this study stems from the frequent use of MT services for translating website content, both by website developers and users. The research is based on the content of monolingual websites of medical and cultural-leisure institutions of the Russian Federation. To identify pre-editing techniques, the study employed the Google Translate system and analyzed the translation of presentational texts into English and German. The analysis revealed errors made by Google Translate when dealing with Russian-English and Russian-German language pairs. The techniques for editing the source text to account for these errors have been proposed.

Key words: pre-editing; presentational text; machine translation; multilingual website; post-editing; source text.

Исследование выполнено при финансовой поддержке Российского научного фонда в рамках научного проекта № 25-28-00223 «Переводоведческое обоснование многоязычия презентационного контента в парадигме ИИ» / The research was supported by Russian Science Foundation (project No. 25-28-00223) “Translatological substantiation of multilinguality of presentational content in the AI paradigm”. (<https://rscf.ru/en/project/25-28-00223/>)

Процесс совершенствования алгоритмов осуществления машинного перевода характеризуется очень высокой скоростью, чем и обусловлена постановка проблем теоретического осмысления и оценки возможностей новейших систем машинного перевода (МП). Несмотря на постоянное развитие данных систем, роль специалиста-переводчика в создании переводного контента является очень важной и зачастую заключается в выполнении постредактирования машинного перевода. На сегодняшний день постредактирование является неотъемлемой составляющей процесса перевода, поскольку возможности современных программ МП позволяют специалистам значительно сократить сроки выполнения перевода, и выполнение последующего редактирования занимает гораздо меньше времени по сравнению с переводом «с нуля».

Однако наряду с распространенной практикой постредактирования актуальным также представляется процесс «предредактирования» (от англ. *pre-editing*), который, согласно своему названию, осуществляется в отличие

от постредактирования перед загрузкой исходного текста в программу машинного перевода. В исследованиях, посвященных автоматизации процесса перевода, отмечается, что предредактирование во многом способствует повышению эффективности и оптимизации перевода в целом и последующего редактирования в частности [1; 2; 3]. Многие переводоведы отмечают, что выполнение предредактирования особенно важно при работе со специальными текстами, насыщенными специфической терминологией и сложным синтаксисом [4; 5]. Кроме того, переводчикам в последнее время зачастую приходится работать с изначально «дефектными» текстами, в которых содержатся ошибки и неточности и которые нуждаются в обязательном предредактировании. Как отмечают У. А. Андрееenkova и О. Г. Шевченко, «причиной этому служат экономия на подготовке документации, привлечение не носителей языка, отсутствие контроля, низкая компетентность авторов, а также массовое использование автоматизированных систем перевода» [6, с. 105–106].

Цель настоящего исследования заключается в определении эффективной технологии предредактирования презентационных текстов, позволяющей учитывать необходимость последующего перевода на иностранные языки уже на этапе создания исходного текста. Актуальность данного исследования обусловлена частотностью практики использования сервисов МП для перевода вербального контента сайтов, как разработчиками сайтов, так и пользователями.

Предредактирование направлено на «улучшение» исходного текста, на повышение его качества, что позволяет сократить или даже полностью исключить необходимость постредактирования. Что же в данном контексте следует понимать под качеством исходного текста? Процесс предредактирования предполагает подробный анализ исходного текста, выполняемый переводчиком или редактором с точки зрения функционирования систем МП и прогнозирования потенциальных ошибок. Таким образом, повысить качество исходного текста в процессе предредактирования означает обеспечить высокую степень его переводимости в программах МП, что, в зависимости от типа и жанра текста, может подразумевать сокращение длины предложений, обеспечение единообразия терминологии, пояснение или опущение культурно-специфичных выражений и т. д.

Актуальным данное направление деятельности переводчиков представляется для работы с многоязычными презентационными материалами, размещаемыми на сайтах различных организаций, что позволяет оптимизировать и унифицировать общий процесс создания переводного контента при минимальном постредактировании. Кроме того, распространенной практикой Интернет-пользователей является также использование систем МП при просмотре страниц иноязычных сайтов, на которых отсутствует версия на языке пользователя, в связи с чем важно обеспечить высокую степень переводимости презентационных гипертекстов в данных программах.

Безусловно для создания «модели» презентационного текста, отличающегося высокой степенью переводимости в системах МП необходимо учитывать характеристики презентационных текстов и ошибки, допускаемые машинными переводчиками. Материалом исследования моделирования и технологии предредактирования презентационных текстов послужил вербальный контент сайтов российских государственных учреждений, включая медицинские учреждения – Волгоградский областной клинический перинатальный центр № 2 (г. Волгоград), Московский многопрофильный научно-клинический центр им. С. П. Боткина (г. Москва), Национальный медицинский исследовательский центр здоровья детей (г. Москва) и их переводы на английский и немецкий языки, выполненные системой машинного перевода Google Translate. Общий объем проанализированного материала составляет около 30 000 п.з. Для анализа использован вербальный контент соответствующих характеристикам презентационного дискурса веб-страниц («О центре», «О перинатальном центре», «О ММНКЦ им. С. П. Боткина» и др.).

При переводе презентационных гипертекстов медицинской тематики, размещенных на сайтах указанных учреждений с русского на английский и немецкий язык с помощью онлайн-сервиса Google Translate были выявлены следующие ошибки. Во-первых, следует отметить вариативность перевода названия медицинского учреждения ММНКЦ им. С. П. Боткина, которая во многом связана с исходной вариативностью формулировок:

Исходный текст [7]	Машинный перевод
ММНКЦ им. С.П. Боткина – крупнейшее многопрофильное лечебное и научно-образовательное учреждение Москвы.	англ. <i>MMNCC named after S.P. Botkin is the largest multidisciplinary medical and scientific-educational institution in Moscow.</i>
	нем. <i>Das S.P. Botkin MMNCC ist die größte multidisziplinäre medizinische und wissenschaftlich-pädagogische Einrichtung in Moskau.</i>
ММНКЦ имени С.П. Боткина – один из 5 опорных центров онкологической помощи в Москве.	англ. <i>SP Botkin MMNCC is one of the 5 reference centers for oncological care in Moscow.</i>
	нем. <i>Das S.P. Botkin MMNCC ist eines der fünf Referenzzentren für onkologische Versorgung in Moskau.</i>
В Боткинской ежегодно принимают порядка 1 миллиона амбулаторных пациентов.	англ. <i>Botkinskaya annually receives about 1 million outpatients.</i>
	нем. <i>Das Botkinskaya-Krankenhaus empfängt jährlich rund eine Million ambulante Patienten.</i>

<i>В 2017 году Боткинская больница признана лучшим многопрофильным стационаром Москвы.</i>	англ. <i>In 2017, Botkinskaya Hospital was recognized as the best multidisciplinary hospital in Moscow.</i>
	нем. <i>2017 wurde das Botkinskaya-Krankenhaus als bestes multidisziplinäres Krankenhaus Moskaus ausgezeichnet.</i>
<i>Каждая восьмая операция в Боткинской относится к категории высокотехнологичной медицинской помощи.</i>	англ. <i>Every eighth operation at Botkin falls into the category of high-tech medical care.</i>
	нем. <i>Jede achte Operation im Botkinskaya-Krankenhaus fällt in die Kategorie der hochtechnologischen medizinischen Versorgung.</i>
<i>С 2019 года правительство Москвы осуществляет в Боткинской больнице масштабную программу комплексного развития.</i>	англ. <i>Since 2019, the Moscow government has been implementing a large-scale comprehensive development program at Botkin Hospital.</i>
	нем. <i>Seit 2019 führt die Moskauer Regierung ein umfassendes Entwicklungsprogramm für das Botkinskaya-Krankenhaus durch.</i>

Как показывают приведенные примеры, многие из переведенных названий на английском и немецком языках являются неадекватными и сложными для восприятия. Официальное название на английском – *Botkin Hospital*, но оно встречается в переводном тексте только один раз. При переводе на немецкий машинный переводчик использовал два варианта – транслитерацию для аббревиатуры ММНКЦ, и имени собственного *Боткинская* и дословный перевод слова *больница*, при этом ни один, ни другой вариант не является правильным, поскольку даже транслитерация была осуществлена по правилам английского языка. В данном случае при создании исходного русскоязычного текста необходимо использовать единообразный вариант названия учреждения (*Боткинская больница*), без употребления аббревиатур (ММНКЦ) и сокращений (*Боткинская*), что позволит значительно снизить вариативность номинации. Удачный пример подобной номинации демонстрируют тексты, размещенные на сайте Национального медицинского исследовательского центра здоровья детей: используется краткое обозначение *Центр*, которое находит соответствующее отражение в переводной версии *‘The Center’ / ‘das Zentrum’*: *В итате Центра – высококвалифицированные специалисты – ‘The Center employs highly qualified specialists’ / ‘Das Zentrum beschäftigt hochqualifizierte Fachkräfte’*; *Все виды исследований в Центре являются общедоступными – ‘All types of research at the Center are publicly available’ / ‘Alle Forschungsarbeiten des Zentrums sind öffentlich zugänglich’*;

Все подразделения Центра имеют современное научное и медицинское оборудование – ‘All departments of the Center have modern scientific and medical equipment’ / ‘Alle Abteilungen des Zentrums verfügen über moderne wissenschaftliche und medizinische Ausstattung’ и др. [8].

Во-вторых, использование сложных синтаксических конструкций и длинных высказываний также создает сложности для восприятия переводного текста. Для презентационных текстов на английском и немецком языках нехарактерны длинные предложения, однако, как показывает анализ переводного текста, Google Translate не членит высказывания, сохраняя исходную конструкцию. Ниже представлены примеры исходного и отредактированного фрагментов оригинального и переводных текстов.

Пример 1

Исходный текст [7]	Машинный перевод
Принимают пациентов Межокружной нефрологический центр (в отреставрированном историческом корпусе № 11), самый красивый и просторный Центр амбулаторной онкологической помощи (ЦАОП) для пациентов Западного округа, хирургический корпус № 7, где размещен Городской центр компетенций по гнойной травма- тологии и хирургии и хирургичес- кий стационар кратковременного пребывания (СКП).	англ. The Interdistrict Nephrology Center (in the restored historical building No. 11), the most beautiful and spacious Outpatient Oncology Care Center (O OCC) for patients of the Western District, Surgical Building No. 7, where the City Competence Center for Purulent Traumatology and Surgery and the Short-Stay Surgical Hospital (SSH) are located, accept patients. нем. Das Interdistrikte Nephrologische Zentrum (im restaurierten historischen Gebäude Nr. 11), das schönste und geräumigste Ambulante Onkologische Versorgungszentrum (O OCC) für Patienten des Westbezirks, das Chirurgiegebäude Nr. 7, in dem sich das Städtische Kompetenzzentrum für eitrige Traumatologie und Chirurgie und das Kurzzeitchirurgische Krankenhaus (SSH) befinden, nimmt Patienten auf.
Отредактированный текст	Машинный перевод
Межокружной нефрологический центр принимает пациентов в отреставрированном истори- ческом корпусе № 11. Это самый красивый и просторный Центр амбулаторной онкологической помощи (ЦАОП) для пациентов Западного округа. Хирургический корпус № 7 также открыт для	англ. The Interdistrict Nephrology Center accepts patients in the restored historical building No. 11. This is the most beautiful and spacious Outpatient Oncology Care Center (O OCC) for patients in the Western District. Surgical building No. 7 is also open to patients. It houses the City Competence Center for Purulent Traumatology and Surgery and the short-stay surgical hospital (SSH).

<i>пациентов. В нем размещен Городской центр компетенций по гнойной травматологии и хирургии и хирургический стационар кратковременного пребывания (СКП).</i>	нем. <i>Das Interdistrikt-Nephrologiezentrum empfängt Patienten im restaurierten historischen Gebäude Nr. 11. Es ist das schönste und geräumigste ambulante onkologische Versorgungszentrum (OOCC) für Patienten im Westbezirk. Auch das chirurgische Gebäude Nr. 7 steht Patienten offen. Es beherbergt das Städtische Kompetenzzentrum für eitrige Traumatologie und Chirurgie sowie das Kurzzeit-Chirurgiekrankenhaus (SSH).</i>

В переводе исходного предложения подлежащее и сказуемое разделены большим количеством дополнительной информации, что нарушает связность и цельность конструкции. В предложенной отредактированной версии вместо одного предложения представлены четыре, что позволяет машинному переводчику создать адекватный и удобный для восприятия перевод на обоих языках.

Пример 2

Исходный текст [9]	Машинный перевод
<i>Основной задачей центра является оказание специализированной медицинской помощи женщинам репродуктивного возраста, беременным с высокой степенью риска развития осложнений в период беременности, родов и послеродовом периоде, роженицам и новорожденным, недоношенным детям, особенно новорожденным с критически низкой массой тела при рождении, родившимся как в перинатальном центре, так и других родовспомогательных учреждениях Волгоградской области и переведенных в него, в том числе оказание высокотехнологичной помощи.</i>	англ. <i>The main objective of the center is to provide specialized medical care to women of reproductive age, pregnant women with a high risk of complications during pregnancy, childbirth and the postpartum period, women in labor and newborns, premature babies, especially newborns with critically low birth weight, born both in the perinatal center and other maternity institutions of the Volgograd region and transferred to it, including the provision of high-tech care.</i> нем. <i>Das Hauptziel des Zentrums besteht darin, Frauen im gebärfähigen Alter, schwangeren Frauen mit einem hohen Risiko von Komplikationen während der Schwangerschaft, der Geburt und im Wochenbett, Frauen in den Wehen und Neugeborenen, Frühgeborenen, insbesondere Neugeborenen mit kritisch niedrigem Geburtsgewicht, die sowohl im Perinatalzentrum als auch in anderen Entbindungseinrichtungen der Region Volgograd geboren und dorthin verlegt werden, eine spezialisierte medizinische Versorgung zu bieten, einschließlich der Bereitstellung einer High-Tech-Versorgung.</i>

Отредактированный текст	Машинный перевод
<i>Основной задачей центра является оказание специализированной высокотехнологичной медицинской помощи <...>.</i>	англ. <i>The main objective of the center is to provide specialized high-tech medical care_...>.</i> нем. <i>Das Hauptziel des Zentrums ist die Bereitstellung spezialisierter medizinischer Hightech-Versorgung ...>.</i>

В примере 2 также наблюдается разделение связанных смысловых элементов большим объемом информации. В начале предложения указано *Основной задачей центра является оказание специализированной медицинской помощи...>*, далее идет перечисление категорий пациентов, которым оказывается помощь, и в самом конце предложения – *<... в том числе оказание высокотехнологичной помощи*. В предлагаемом отредактированном варианте фраза *<...в том числе оказание высокотехнологичной помощи* встроена в начало предложения путем использования прилагательного: *Основной задачей центра является оказание специализированной высокотехнологичной медицинской помощи...>*, что позволяет добиться адекватного перевода на английский и немецкий языки, который будет намного легче восприниматься получателями.

Пример 3

Исходный текст [8]	Машинный перевод
<i>Центр объединяет детских врачей страны для совместной клинической, научной и учебной деятельности: ежегодно проводятся конгрессы педиатров России, тематические научно-практические конференции, школы для специалистов.</i>	англ. <i>The center unites pediatricians of the country for joint clinical, scientific and educational activities: congresses of pediatricians of Russia, thematic scientific and practical conferences, schools for specialists are held annually.</i> нем. <i>Das Zentrum vereint Kinderärzte des Landes für gemeinsame klinische, wissenschaftliche und pädagogische Aktivitäten: Jährlich finden Kongresse der Kinderärzte Russlands, thematische wissenschaftliche und praktische Konferenzen sowie Schulen für Spezialisten statt.</i>

Отредактированный текст	Машинный перевод
<i>Центр объединяет детских врачей страны для совместной клинической, научной и учебной деятельности. Ежегодно проводятся конгрессы педиатров России, тематические научно-практические конференции, школы для специалистов.</i>	англ. <i>The center unites pediatricians of the country for joint clinical, scientific and educational activities. Congresses of pediatricians of Russia, thematic scientific and practical conferences, schools for specialists are held annually.</i> нем. <i>Das Zentrum vereint Kinderärzte des Landes für gemeinsame klinische, wissenschaftliche und pädagogische Aktivitäten. Jährlich finden Kongresse russischer Kinderärzte, thematische wissenschaftliche und praktische Konferenzen sowie Schulungen für Fachkräfte statt.</i>

В примере 3 также показано изменение структуры исходного предложения в процессе предредактирования путем членения на более короткие синтагмы.

Стоит обратить внимание на использование в исходном тексте аббревиатур, характерных для медицинского дискурса. В тексте о ММНКЦ им. С. П. Боткина их достаточно много, и соответственно, все они сохранены машинным переводчиком: *высокотехнологичная медицинская помощь (ВМП)* – ‘*high-tech medical care (HMC)*’ / ‘*Hightech-Medizinische Versorgung (HMC)*’; *Центр амбулаторной онкологической помощи (ЦАОП)* – ‘*the Center for Outpatient Oncological Care (CAOC)*’ / ‘*Zentrum für Ambulante Onkologische Versorgung (CAOC)*’; *стационар кратковременного пребывания (СКП)* – ‘*Short-Stay Surgical Hospital (SSH)*’ / ‘*Kurzzeitklinik (SSH)*’; *Медицинский симуляционный центр (МЦЦ)* – ‘*Medical Simulation Center (MSC)*’ / ‘*Medizinisches Simulationszentrum (MSC)*’ [7]. Учитывая тот факт, что многие аббревиатуры являются локальными и образованы от переводных калек с русского языка, то в переводе в них нет необходимости, они перегружают текст, и в процессе предредактирования текста для машинного перевода их необходимо опустить. Кроме того, при переводе данных аббревиатур на немецкий язык Google использует англоязычный вариант, который не совпадает с первыми буквами компонентов, входящих в состав полных наименований на немецком языке, что является абсолютно некорректным и недопустимым без дополнительных пояснений и может служить доказательством тому, что в данном случае система не переводит напрямую, а привлекает английский как язык посредник.

Среди других параметров, которые следует учитывать при предредактировании или составлении презентационных текстов, которые будут переводиться системами машинного перевода на английский и/или немецкий языки,

можно выделить также следующие: грамматическая корректность (четкая структура предложений, минимизация деепричастных и длинных причастных оборотов, правильное согласование рода, числа и падежей и т. д.); конкретность и четкость (например, избегание местоимений с неясной референцией, учет наличия артиклей в английском и немецком языках); отказ от морфологической сложности (например, длинные цепочки родительного падежа, сложные глагольные формы); правильная пунктуация.

Таким образом, на основе проведенного анализа можно сделать вывод о языковых характеристиках, лежащих в основе модели презентационного текста на русском языке, отличающегося высокой степенью переводимости в системах машинного перевода. К таким характеристикам относится использование коротких предложений, единообразие лексической номинации, отсутствие аббревиатур, грамматическая и пунктуационная корректность. Решение научно-исследовательской проблемы разработки принципов предредактирования текста для машинного перевода и моделирования презентационного текста позволяет оптимизировать процессы создания многоязычного контента с использованием систем МП в рамках Интернет-дискурса и повысить его качество.

ЛИТЕРАТУРА

1. Hiraoka Y., Yamada M. Pre-editing plus neural machine translation for subtitling: effective pre-editing rules for subtitling of TED Talks // Proceedings of MT. Dublin MT Summit XVII. 2019. Vol. 2. P. 64–74.
2. Мэнсюе Ч. Влияние предредактирования на эффективность машинного перевода исторической литературы на примере перевода "Истории науки и техники Китая" // Вопросы истории. 2023. № 9–1. С. 198–207. DOI: 10.31166/VoprosyIstorii202309Statyi26.
3. Taufik A. Pre-editing of Google neural machine translation [Electronic resource] // Journal of English Language and Culture. 2020. Vol. 10. № 2. P. 64–74. DOI: <http://dx.doi.org/10.30813/jelc.v10i2.2137>.
4. Животова А. А., Бердонос В. Д. Оптимизационное предредактирование узкоспециальных русскоязычных текстов для их машинного перевода на английский язык // Информационные и математические технологии в науке и управлении. 2024. № 2 (34). С. 169–182. DOI: 10.25729/ESI.2024.34.2.016.
5. Коканова Е. С., Пак Н. С. Предредактирование текстов для машинного перевода как метод исследования английских конструкций с причастием в медицинском дискурсе // Когнитивные исследования языка. 2022. № 3 (50). С. 676–680.
6. Андрееenkova У. А., Шевченко О. Г. Влияние аномалий дефектных текстов на машинный перевод и предредактирование дефектных текстов (на материале англоязычных инструкций) // Социальные и гуманитарные науки в условиях вызовов современности. Материалы III Всероссийской научно-практической конференции молодых ученых с международным участием. В 2-х частях. Комсомольск-на-Амуре, 2025. С. 105–107.

7. ММНКЦ им. С. П. Боткина – Московский многопрофильный научно-клинический центр им. С. П. Боткина [Электронный ресурс]. URL: <https://botkinmoscow.ru/about/o-bolnitse/botkinskaya-bolnitsa-segodnya/> (дата обращения: 03.07.2025).
8. НМИЦ здоровья детей – Национальный медицинский исследовательский центр здоровья детей [Электронный ресурс]. URL: <https://nczd.ru/about/> (дата обращения: 03.07.2025).
9. ВОКПЦ № 2 – Волгоградский областной клинический перинатальный центр № 2 [Электронный ресурс]. URL: <https://vokpc2.ru/about-vokpc2/>. (дата обращения: 03.07.2025).

УДК 81`322.4

Епимахова Александра Сергеевна
кандидат филологических наук,
доцент базовой кафедры технологий
и автоматизации перевода в бюро
переводов «АКМ-Вест»
Северный (Арктический)
федеральный университет имени
М. В. Ломоносова
г. Архангельск, Россия

Aleksandra Epimakhova
PhD in Philology,
Associate Professor of the Department
of Translation Technology and Practice
at AKM-West
Northern (Arctic) Federal University
Arkhangelsk, Russia
a.epimahova@narfu.ru

Коканова Елена Сергеевна
кандидат филологических наук,
заведующий базовой кафедрой
технологий и автоматизации перевода
в бюро переводов «АКМ-Вест»
Северный (Арктический)
федеральный университет
имени М.В. Ломоносова
г. Архангельск, Россия

Elena Kokanova
PhD in Philology,
Head of the Department of Translation
Technology and Practice at AKM-West
Northern (Arctic) Federal University
Arkhangelsk, Russia
e.s.kokanova@narfu.ru

АВТОМАТИЧЕСКАЯ ОЦЕНКА КАЧЕСТВА МАШИННОГО ПЕРЕВОДА В ЯЗЫКОВОЙ ПАРЕ С МАЛОРЕСУРСНЫМ ЯЗЫКОМ

Использование метрик автоматической оценки качества стало неотъемлемой частью разработки систем машинного перевода. Выбор метрик определяется такими факторами, как языковая пара и направление перевода, возможность интуитивной интерпретации полученных результатов, популярность метрики среди разработчиков и исследователей. В статье проводится экспериментальное исследование по разработке модуля оценки качества машинного перевода для языковой пары ненецкий-русский. В результате исследования для автоматической оценки качества выбраны классические метрики BLEU, sacreBLEU, TER, CharacTER, RIBES, chrF++ для обоих направлений перевода и нейросетевая метрика BERTScore только для направления ненецкий-русский. Ограничения связаны с тем, что необходимо использовать метрики, либо рассчитываемые независимо от конкретного языка, либо поддерживающие данный язык. Число метрик, поддерживающих русский язык, ограничено. Ненецкий язык относится к малоресурсным и не поддерживается метриками. Использование нескольких метрик позволяет получить более комплексную оценку.

Ключевые слова: машинный перевод; автоматическая оценка качества машинного перевода; метрики автоматической оценки качества машинного перевода; малоресурсный язык; ненецкий язык; языковая пара ненецкий-русский.

AUTOMATIC MACHINE TRANSLATION EVALUATION FOR A LOW-RESOURCE LANGUAGE

Automatic evaluation metrics are an integral part of machine translation systems development. The choice of metrics is determined by the source and target languages, possibility of intuitive interpretation of the scores, and the popularity of the metric among developers and

researchers. The paper presents an experimental study on the development of a machine translation evaluation module for the Nenets-Russian language pair. As a result of the study, the traditional metrics BLEU, sacreBLEU, TER, CharacTER, RIBES, chrF++ were selected for automatic evaluation in both translation directions. BERTScore neural network metric was selected only for the Nenets-Russian direction. The chosen metrics should either be calculated independently of a specific language, or support the given languages. However, the number of metrics supporting Russian is limited. As for the Nenets language, it is low-resourced and therefore not supported by any metric. Using multiple metrics permits to get a more comprehensive assessment.

Key words: machine translation; automatic machine translation evaluation; automatic machine translation evaluation metrics; low-resource language; Nenets language; Nenets-Russian language pair.

В настоящее время автоматическая оценка качества машинного перевода (МП) с использованием метрик стала не просто альтернативой экспертной оценке, призванной снизить временные и финансовые затраты. Она включена в цикл разработки систем МП и позволяет в числовой форме выразить оценку качества МП (предполагается, что эта оценка соотносится с экспертной оценкой) [1]. Кроме того, автоматическая оценка позволяет сопоставлять системы МП между собой, определять, какая из них покажет лучший результат для определенного типа текста [2, р. 7297]. Важными требованиями к метрикам автоматической оценки качества МП, помимо возможности различать по качеству разные системы МП, а также корреляции с результатами экспертной оценки МП, являются возможность интуитивной интерпретации полученных результатов, ясность и применимость к разным языкам перевода [3]. Подключение экспертной оценки позволяет подкрепить автоматическую оценку [4, с. 108].

Выбор метрик важен при разработке и использовании систем МП, а также публикации полученных результатов [5], так как разные метрики позволяют не только учесть особенности языковой пары, но и сделать результаты более доступными благодаря использованию распространенных метрик. При выборе метрик автоматической оценки МП учитываются разные факторы.

Факторы, определяющие выбор метрик автоматической оценки МП

При выборе метрики важным фактором является возможность ее расчета с использованием имеющихся в открытом доступе библиотек, таких как NLTK, sacrebleu, evaluate. Из множества разработанных метрик лишь самые популярные реализованы в библиотеках. Они позволяют с использованием языка программирования Python загрузить на вход текст или предложение, передать параметры оценки и получить результат.

Результаты оценки по одной и той же метрике могут иметь разное числовое выражение. Например, результаты метрики BLEU (BiLingual Evaluation Understudy [6]), рассчитываемой на основе пересечения n-грамм, зависят от предварительной подготовки текста (она может включать удаление знаков препинания, приведение к нижнему регистру и т. д.), используемого механизма токенизации, количества предложений-эталонов и длины

n-грамм. Поскольку эти параметры не всегда представлены в публикациях, следует с осторожностью сравнивать оценки BLEU из разных статей [7]. При этом в 2021 году BLEU использовалась в 98,8 % публикаций, причем в 74,3 % случаев без привлечения других метрик [2, р. 7299]. Для стандартизации BLEU была создана SacreBLEU, имеющая встроенный токенизатор и использующая установленные параметры оценки (для поддерживаемых языков возможно подключение данных из набора экспертной оценки WMT, Workshop on Machine Translation) [8].

Метрики, для которых есть возможность расчета с использованием библиотек, могут поддерживать не все языки. Например, на основе BLEU построена метрика NIST [9], особенность которой состоит в том, что каждой n-грамме в МП и в эталонном переводе присваивается определенный вес, зависящий от частности элемента в языке перевода, и этот вес учитывается при расчете метрики. Метрика NIST, рассчитываемая с использованием библиотеки NLTK, реализована для английского языка. Метрика METEOR (Metric for Evaluation of Translation with Explicit ORdering) [10] также имеет в основе пересечение n-грамм. Ее ключевое отличие от BLEU состоит в том, что учитываются не только совпадения полных слов, но также наличие однокоренных слов и синонимов. Поэтому метрика METEOR реализована в библиотеке NLTK с подключением по умолчанию тезауруса WordNet [11], т. е. имеет смысл использовать ее только если языком перевода является английский. Следовательно, преимущество BLEU состоит в ее универсальности и удобстве использования для разных языков.

Среди классических метрик, которые не требуют поддержки конкретного языка, следует назвать RIBES, ROUGE, TER, CharacTER, chrF++.

RIBES (Rank-based Intuitive Bilingual Evaluation Score) [12] позволяет учесть позиции слов в предложении, что важно для оценки МП, когда порядок слов в языках сильно различается.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [13] изначально разработана для оценивания результата суммаризации текста (сопоставление аннотации с эталонной аннотацией) и позволяет оценить количество элементов эталонного текста, представленных в оцениваемом.

Метрика TER (Translation Edit Rate) [14] основана на подсчете минимального количества правок, которое необходимо для того, чтобы получить перевод, наиболее близкий к эталонному (итоговый показатель – отношение числа требуемых правок к длине предложения, усредненной по всем эталонам). CharacTER (Translation Edit Rate on Character Level) [15] оценивает необходимое количество правки, но рассчитывается его на основе символов, а не слов, как TER (что актуально для языков с богатой морфологией, так как иногда требуется исправление не целого слова, а только его части).

В основе метрики chrF++ (correlations with human rankings for 6-gram F1-score) [16, р. 612] лежит подсчет совпадений последовательно идущих символов, а также совпадения униграмм и биграмм.

Исследования показывают, что классические метрики уступают в корреляции с экспертной оценкой нейросетевым [17], хотя результаты оценки последних труднее поддаются интерпретации, так как логика ее получения остается неизвестной [18]. Однако нейросетевые метрики имеют в основе модели, обученные на данных для конкретных языков, поэтому они реализованы не для всех языков (прежде всего для английского).

Русский язык поддерживает метрика BERTScore [19]. BERTScore вычисляет близость двух предложений как сумму косинусных сходств между эмбедингами слов в предложениях. BERTScore выдает комплекс из трех оценок: Precision (точность), Recall (полнота) и F1 (F-мера). Precision показывает, насколько для каждого токена МП есть близкий ему токен эталонного предложения; Recall показывает, насколько для каждого токена эталонного предложения есть близкий ему токен МП. F1 учитывает оба эти параметра и рассчитывается по формуле: $F1 = 2 * (Precision * Recall) / (Precision + Recall)$. BERTScore использует 12 предобученных моделей эмбедингов для разных языков и обучена на корпусе WMT18 metric evaluation dataset, который содержит МП 149 разных систем для 14 языковых пар, эталонные варианты перевода, а также результаты экспертной оценки WMT [19, p. 5].

Также русский язык поддерживает COMET (Crosslingual Optimized Metric for Evaluation of Translation) [20]. COMET – это нейросетевая метрика, построенная на PyTorch. При обучении модели использовались не только языковые данные, но и экспертные оценки переводов WMT. Особенность COMET состоит в том, что она получает на вход МП, эталонный перевод и текст оригинала. Следовательно, возможность применения метрики ограничена поддержкой обоих языков языковой пары.

Несмотря на более высокую точность нейросетевых метрик, преимущество традиционных метрик состоит в том, что многие из них не требуют привязки к конкретному языку. Это тем более важно, если в языковую пару входит малоресурсный язык, т. е. язык «с малым объемом электронных ресурсов, доступных для обработки» [21, с. 679]). Малоресурсные языки зачастую не имеют корпусов, словарей, электронных библиотек для обработки естественного языка.

Экспериментальное исследование

В рамках эксперимента по разработке системы МП для языковой пары «ненецкий-русский» были выбраны метрики, которые могут быть использованы для оценки качества МП в обоих направлениях.

При выборе учитываются следующие критерии: (1) расчет с использованием библиотек; (2) применимость к любому языку или поддержка языка перевода; (3) использование только эталонного перевода, без текста оригинала (в направлении ненецкий-русский); (4) оригинальность лежащих в основе критериев; (5) использование в публикациях. В результате собран набор метрик, учитывающих разные аспекты (совпадение n-грамм, объем правки, порядок слов и др.).

В результате из представленных выше метрик для направления перевода «ненецкий-русский» выбраны классические (BLEU, sacreBLEU, TER, CharacTER, RIBES, chrF++) и нейросетевая BERTScore. Они реализованы в библиотеках NLTK, evaluate и sacrebleu. COMET не может быть использована, так как она поддерживает русский язык, но не поддерживает ненецкий, требуя загрузки текста оригинала. Было принято решение исключить из набора метрик ROUGE, так как она изначально не предназначена для оценки МП; редко используется в современных публикациях, посвященных МП; выдает громоздкую цепочку оценок; при этом схожие параметры оцениваются chrF++. BLEU и sacreBLEU, TER и CharacTER по сути являются вариантами одной метрики, однако они часто используются в публикациях, поэтому целесообразно иметь оба варианта оценок. RIBES отличается от прочих метрик таким уникальным параметром как учет порядка слов, что важно для ненецкого и русского языков, имеющих разный синтаксис.

Для направления перевода «русский-ненецкий» могут быть использованы только классические метрики (выбраны BLEU, sacreBLEU, TER, CharacTER, RIBES, chrF++).

Для разработки алгоритма автоматической оценки (на уровне предложения) использован набор, содержащий 10 предложений на русском языке, которые имеют одно предложение-эталон. Оцениваемые предложения моделируют МП и представляют собой варианты перевода исходного предложения, в том числе содержащие ошибки.

Исходное предложение: *Хувы ялумдаукауна нисяв си''ми сиде.*

Эталонный перевод: *Отец разбудил меня на утренней заре.*

Оцениваемые предложения:

1. *Отец разбудил меня на утренней заре.*
2. *На утренней заре меня разбудил отец.*
3. *Папа разбудил меня на утренней заре.*
4. *Мой отец разбудил меня на утренней заре.*
5. *Отец разбудил меня утром на заре.*
6. *Отец разбудил меня на заре.*
7. *Отец не разбудил меня на утренней заре.*
8. *Мать разбудила меня на утренней заре.*
9. *У меня есть летний чум.*
10. *У нас есть летний чум.*

Предложения, выступающие в роли МП, имеют характеристики, которые позволяют увидеть особенности разных метрик: полное совпадение с эталоном (1), изменение порядка слов (2), замена знаменательного слова на синоним (3), добавление служебного слова, которое не меняет смысл предложения (4), использование однокоренного слова (5), удаление слова без изменения смысла предложения (6), добавление отрицания (7), замена слова на другое, используемое в схожих контекстах, с изменением смысла (8), совпадение только одного слова (9), полное несовпадение (10). В последних двух случаях перевод полностью неверный.

Предобработка предложений состоит в приведении к нижнему регистру и удалении знаков препинания, которые иначе могут быть оценены как совпадающие токены. Если для расчета метрики требуется предварительная токенизация, она производится с использованием метода `split()`.

Таблица 1

Результаты оценки с использованием традиционных метрик

	Предложение	BLEU (1,2,3,4)	Sacre BLEU	RIBES	TER	charac TER	chr++
1	Отец разбудил меня на утренней заре.	1.0, 1.0, 1.0, 1.0	100.0	1.0	0.0	0.0	100.0
2	На утренней заре меня разбудил отец.	1.0, 0.63, 0.47, 0.0	82.78	0.2	50.0	0.36	72.79
3	Папа разбудил меня на утренней заре.	0.83, 0.82, 0.79, 0.76	85.94	0.96	16.67	0.11	85.4
4	Мой отец разбудил меня на утренней заре.	0.86, 0.85, 0.83, 0.81	90.46	0.96	16.67	0.1	97.86
5	Отец разбудил меня утром на заре.	0.83, 0.58, 0.44, 0.0	66.04	0.57	33.33	0.31	61.06
6	Отец разбудил меня на заре.	0.82, 0.71, 0.65, 0.58	64.07	0.59	16.67	0.35	65.77
7	Отец не разбудил меня на утренней заре.	0.86, 0.76, 0.7, 0.64	88.28	0.96	16.67	0.08	89.85
8	Мать разбудила меня на утренней заре.	0.67, 0.63, 0.59, 0.51	78.39	0.9	33.33	0.14	75.89
9	У меня есть летний чум.	0.16, 0.0, 0.0, 0.0	9.44	0.0	83.33	1.0	12.31
10	У нас есть летний чум.	0, 0, 0, 0	3.4	0.0	100.0	1.0	8.54

BLEU рассчитывается на 4 уровнях (от униграмм до 4-грамм) с использованием библиотеки NLTK, из которой импортируется `sentence_bleu`, т. е. оценка на уровне предложения, а не текста. Как показывает таблица 1, получены оценки от максимальной для первого предложения до нуля для последнего. Самые высокие оценки получены на уровне униграмм, при этом перестановка слов вовсе не отразилась на оценке. Данная метрика оценивает пересечения n-грамм, поэтому она не может полностью отразить качество перевода. В частности, при добавлении отрицания (7) оценка оказалась выше, чем при изменениях, не повлиявших на смысл: удалении прилагательного, значение которого понятно из контекста (6), замене существительного на синоним (3).

Интересно, что для последнего предложения оценка SacreBLEU оказалась не равна нулю, хотя она и очень низкая. Вероятно, это связано с особенностями выравнивания, так как BLEU и SacreBLEU предназначены изначально для оценки целого текста, а не отдельных предложений, и SacreBLEU не имеет отдельной реализации для оценки предложений.

Что касается оценки метрики RIBES, она резко падает на предложениях с измененным порядком слов. В данном случае для русского языка оно влияет на тема-рематическое членение предложения, но не меняет фактическую информацию. Оценка RIBES будет важна для направления перевода русский-ненецкий, так как ненецкий язык имеет фиксированный порядок слов.

Для TER и characTER, в отличие от остальных метрик, минимальное значение желательнее максимального (они рассчитывают количество вносимой в МП правки). Если привести значения TER и characTER к общему знаменателю и сопоставить их, видно, что оценки различаются между собой, так как TER ведет расчет по словам, а characTER – по символам.

Поскольку chrF++ учитывает присутствие единиц эталонного перевода в МП, наиболее высокую оценку (кроме полного совпадения) получает предложение (4), куда было добавлено притяжательное местоимение.

Последняя из рассматриваемых метрик – BERTScore.

Таблица 2

Результаты оценки с использованием BERTScore

	Предложение	Precision	Recall	F1
1	Отец разбудил меня на утренней заре.	1.0	1.0	1.0
2	На утренней заре меня разбудил отец.	0.9	0.9	0.9
3	Папа разбудил меня на утренней заре.	0.98	0.98	0.98
4	Мой отец разбудил меня на утренней заре.	0.94	0.98	0.96
5	Отец разбудил меня утром на заре.	0.94	0.9	0.92
6	Отец разбудил меня на заре.	0.97	0.88	0.92
7	Отец не разбудил меня на утренней заре.	0.97	0.99	0.98
8	Мать разбудила меня на утренней заре.	0.97	0.98	0.98
9	У меня есть летний чум.	0.71	0.67	0.69
10	У нас есть летний чум.	0.71	0.67	0.69

Интересно, что значение метрики BERTScore ни для одного из предложений не равно нулю, хотя последнее предложение не имеет ничего общего с эталонным. Также обращает на себя внимание тот факт, что высокую оценку получили предложения 7 и 8, содержащие фактические ошибки.

Для оценки МП в направлении «русский-ненецкий» BERTScore использоваться не может, так как необходима поддержка языка перевода.

Использование набора метрик позволяет не только оценить МП с разных позиций, но и привлечь для анализа оценки, полученные для языковых пар, включающих другие языки коренных народов России. Например, для оценки МП в языковой паре «татарский-русский» для оценки результатов используется BLEU [22], а в языковой паре «эрзянский-русский» – BLEU и ChrF++ [23].

Отметим, что время расчета для метрик различается. Поскольку время может зависеть от таких факторов, как загруженность системы, код был запущен 5 раз для расчета среднего показателя. Показатели могут измениться, если код будет запускаться более 10 раз и на разных графических процессорах.

Таблица 3

Время расчета метрики для 10 предложений

Метрика	Время расчета, сек.					Среднее время, сек
BLEU	0,005	0,010	0,006	0,007	0,005	0,006
SacreBLEU	0,077	0,090	0,040	0,063	0,004	0,054
RIBES	0,008	0,004	0,004	0,005	0,004	0,005
TER	0,045	0,039	0,065	0,058	0,041	0,005
characTER	0,078	0,032	0,033	0,037	0,050	0,046
chrF++	0,156	0,039	0,036	0,053	0,038	0,064
BERTScore	3,846	25,310	2,664	4,532	2,359	7,742

Как показывает таблица 3, чем больше параметров учитывает метрика, тем больше времени требуется на вычисление. Нейросетевая метрика BERTScore требует для расчета результатов гораздо больше времени и ресурсов, по сравнению с использованием классических метрик.

Представленный набор из 10 предложений моделирует МП, позволяя определить, как на оценки метрик влияют разные отклонения от эталонного перевода (варианты и ошибки). В дальнейшем для оценки результатов МП используется набор реальных предложений, представляющих разные типы текста в наборе данных [24]. Предложения, включенные в этот контрольный набор, удаляются из набора данных перед обучением модели МП.

Использование метрик автоматической оценки качества МП позволяет быстро получить непредвзятую оценку для контрольного набора предложений. Целесообразность использования набора метрик определяется направлением перевода для малоресурсных языков. Кроме того, использование метрик «золотого стандарта» позволяет вписать исследование в систему публикаций связанной тематики.

Особенность исследования малоресурсных языков состоит в том, что для оценки качества МП могут использоваться только классические метрики, которые не требуют поддержки конкретного языка. Разработка нейросетевых метрик для малоресурсных языков остается делом ближайшего будущего, поскольку на настоящем этапе для многих из них стоит задача сбора данных. Параллельно будет происходить совершенствование систем МП.

ЛИТЕРАТУРА

1. Babych B. Automated MT evaluation metrics and their limitations [Electronic resource] // Revista Tradumàtica: tecnologies de la traducció. 2014. № 12, Traducció i qualitat. P. 464–470. URL: <https://www.raco.cat/index.php/Tradumatica/article/download/286868/375090> (date of access: 14.07.2025).
2. Marie B., Fujita A., Rubino R. Scientific Credibility of Machine Translation Research: A Meta-Evaluation of 769 Papers [Electronic resource] // Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021. P. 7297–7306. DOI: <https://doi.org/10.18653/v1/2021.acl-long.566>.
3. The NIST 2008 Metrics for machine translation challenge—overview, methodology, metrics, and results [Electronic resource] / M. Przybocki, K. Peterson, S. Bronsart, G. Sanders // Machine Translation. 2009. № 23(2/3). P. 71–103. DOI: <https://doi.org/10.1007/s10590-009-9065-6>.
4. Нуриев В. А., Егорова А. Ю. Методы оценки качества машинного перевода: современное состояние // Информатика и ее применение. 2021. Том 15. Выпуск 2. С. 104–111.
5. Берендяев М. В., Гилин М. И., Коканова Е. С. Генеративный искусственный интеллект и оценка качества перевода // Человек: Образ и сущность. 2025. №2(62). [в печати]
6. BLEU: a method for automatic evaluation of machine translation [Electronic resource] / K. Papineni, S. Roukos, T. Ward, W.-J. Zhu // Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02). Association for Computational Linguistics, USA, 2002. P. 311–318. DOI: <https://doi.org/10.3115/1073083.1073135>.
7. Коканова Е. С., Поротова П. В. Автоматическая метрика оценка качества машинного перевода BLEU // Профессия переводчика: вызовы и перспективы: Сборник студенческих научных статей по материалам работы III Всероссийской молодежной научной онлайн-конференции «Современные техно-

логии в переводе. Машинный перевод: от стереотипов к новым возможностям», Архангельск, 24 ноября 2022 года. Архангельск: Северный (Арктический) федеральный университет им. М. В. Ломоносова, 2023. С. 44–47.

8. Post M. A Call for Clarity in Reporting BLEU Scores [Electronic resource] // Proceedings of the Third Conference on Machine Translation: Research Papers. 2018. P. 186–191. DOI: <https://doi.org/10.18653/v1/W18-6319>.

9. Doddington G. Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics [Electronic resource] // HLT '02: Proceedings of the second international conference on Human Language Technology Research. 2002. P. 138–145. URL: <https://dl.acm.org/doi/10.5555/1289189.1289273> (date of access: 14.07.2025).

10. Banerjee S., Lavie A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments [Electronic resource] // Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. 2005. P. 65–72. URL: <https://aclanthology.org/W05-0909/> (date of access: 14.07.2025).

11. WordNet [Electronic resource]. URL: <https://wordnet.princeton.edu/> (date of access: 14.07.2025).

12. Automatic Evaluation of Translation Quality for Distant Language Pairs [Electronic resource] / H. Isozaki, T. Hirao, K. Duh, K. Sudoh, H. Tsukada // Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. 2010. P. 944–952. URL: <https://aclanthology.org/D10-1092/> (date of access: 14.07.2025).

13. Lin C.-Y. ROUGE: A Package for Automatic Evaluation of Summaries [Electronic resource] // Annual Meeting of the Association for Computational Linguistics. 2004. URL: <https://aclanthology.org/W04-1013/> (date of access: 14.07.2025).

14. Study of Translation Edit Rate with Targeted Human Annotation [Electronic resource] / M. Snover, B. Dorr, R. Schwartz, L. Micciulla, J. Makhoul // Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers. 2006. P. 223–231. URL: <https://aclanthology.org/2006.amta-papers.25/> (date of access: 14.07.2025).

15. CharacTer: Translation Edit Rate on Character Level [Electronic resource] / W. Wang, J.-T. Peter, H. Rosendahl, H. Ney // Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers. 2016. P. 505–510. DOI: <https://doi.org/10.18653/v1/W16-2342>.

16. Popović M. CharacTer: chrF++: words helping character n-grams [Electronic resource] // Proceedings of the Second Conference on Machine Translation. 2016. P. 612–618. DOI: <https://doi.org/10.18653/v1/W17-4770>.

17. Results of WMT22 Metrics Shared Task: Stop Using BLEU – Neural Metrics Are Better and More Robust [Electronic resource] / M. Freitag, R. Rei, N. Mathur, C.-k. Lo, C. Stewart, E. Avramidis, T. Kocmi, G. Foster, A. Lavie, A. F. T. Martins // Proceedings of the Seventh Conference on Machine Translation (WMT). 2022. P. 46–68. URL: <https://aclanthology.org/2022.wmt-1.2/> (date of access: 14.07.2025).

18. COMET: A Neural Framework for MT Evaluation [Electronic resource] / R. Rei, C. Stewart, A. C. Farinha, A. Lavie // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020. P. 2685–2702. DOI: <https://doi.org/10.18653/v1/2020.emnlp-main.213>.
19. BERTScore: Evaluating Text Generation with BERT [Electronic resource] / T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, Y. Artzi. 2020. DOI: <https://doi.org/10.48550/arXiv.1904.09675>.
20. The Inside Story: Towards Better Understanding of Machine Translation Neural Evaluation Metrics [Electronic resource] / R. Rei, N. M. Guerreiro, M. Treviso, L. Coheur, A. Lavie, A. F. T. Martins // Proceedings of the Seventh Conference on Machine Translation (WMT). 2023. P. 46–68. DOI: <https://doi.org/10.18653/v1/2023.acl-short.94>
21. Кипяткова И. С., Кагиров И. А., Долгушин М. Д. Применение предварительно обученных многоязычных моделей для распознавания карельской речи [Электронный ресурс] // Информатика и автоматизация. 2025. № 24 (2). С. 604-630. DOI: <https://doi.org/10.15622/ia.24.2.9>.
22. Application of Low-resource Machine Translation Techniques to Russian-Tatar Language Pair [Electronic resource] / A. Valeev, I. Gibadullin, A. Khusainova, A. Khan. 2019. DOI: <https://doi.org/10.48550/arXiv.1910.00368>.
23. Dale D. The first neural machine translation system for the Erzya language [Electronic resource] // Proceedings of the first workshop on NLP applications to field linguistics. 2022. P. 45–53. URL: <https://aclanthology.org/2022.fieldmatters-1.6/> (date of access: 14.07.2025).
24. Епимахова А. С., Коканова Е. С. Тундровый ненецкий язык: несбалансированность набора данных // СЛОВА и ЦИФРЫ: материалы I Международной научно-практической конференции. Йошкар-Ола – Москва, 4–5 апреля 2025 г. / Министерство науки и высшего образования Российской Федерации, Поволжский государственный технологический университет, Университет науки и технологий МИСИС. Йошкар-Ола: Поволжский государственный технологический университет, 2025. С. 110–115.

УДК 811.134.2: 372.881.1

Мартынова Татьяна Константиновна
студент докторантуры
Университет Гранады
г. Гранада, Испания

Tatyana Martynova
PhD Student
University of Granada
Granada, Spain
martynovatk@correo.ugr.es

МАШИННЫЙ ПЕРЕВОД И ЛЕКСИКОГРАФИЧЕСКАЯ ПОДДЕРЖКА: СРАВНИТЕЛЬНЫЙ ПОДХОД В УСЛОВИЯХ ЦИФРОВОЙ ТРАНСФОРМАЦИИ ЯЗЫКОВОГО ОБРАЗОВАНИЯ

В статье рассматриваются возможности и ограничения использования современных систем машинного перевода и цифровых лексикографических ресурсов при обучении переводу художественного текста в контексте цифровизации языкового образования. Эмпирической базой исследования послужили фрагменты произведений современной русской прозы, насыщенные культурно специфической лексикой, идиоматикой и стилистически маркированными конструкциями. Анализ выполнен на материале трех автоматизированных переводческих платформ – Google Translate, Yandex Translate и ChatGPT – с привлечением многоуровневой методики оценки, включающей орфографический, лексический, морфологический, грамматический, синтаксический и прагматический уровни. Полученные результаты свидетельствуют о том, что при наличии лексикографической поддержки и педагогической интеграции машинный перевод может эффективно использоваться в образовательных целях как вспомогательное средство. Вместе с тем его применение без критической интерпретации сопряжено с рисками искаженной передачи авторского замысла и культурных реалий. В заключение подчеркивается целесообразность включения в программы подготовки переводчиков модулей, ориентированных на формирование навыков критического анализа и работы с цифровыми инструментами в сфере художественного перевода.

Ключевые слова: машинный перевод; лексикографическая компетенция; языковое образование; лингводидактика; переводческая подготовка.

MACHINE TRANSLATION AND LEXICOGRAPHIC SUPPORT: A COMPARATIVE APPROACH IN THE CONTEXT OF THE DIGITAL TRANSFORMATION OF LANGUAGE EDUCATION

The article explores the possibilities and limitations of using modern machine translation systems and digital lexicographic resources in teaching the translation of literary texts within the context of language education digitalization. The empirical basis of the study consists of excerpts from contemporary Russian prose, rich in culturally specific vocabulary, idiomatic expressions, and stylistically marked constructions. The analysis was conducted using three automated translation platforms – Google Translate, Yandex Translate, and ChatGPT – and involved a multi-level evaluation methodology encompassing orthographic, lexical, morphological, grammatical, syntactic, and pragmatic dimensions. The findings suggest that, when supported by lexicographic tools and pedagogical integration, machine translation can be effectively employed in educational settings as an auxiliary aid. However, its use without critical

interpretation entails the risk of distorting the author's intent and misrepresenting cultural references. The article concludes by emphasizing the importance of incorporating modules aimed at developing skills in critical analysis and digital tool usage for literary translation into translator training programs.

Key words: machine translation; lexicographic competence; language education; language didactics; translator training.

В условиях стремительного внедрения цифровых технологий в сферу образования усиливается интерес к использованию автоматизированных инструментов в подготовке переводчиков – специалистов в области межкультурной коммуникации [1; 2; 3; 4; 5]. Все чаще в фокус научных исследований попадают как преимущества и ограничения машинного перевода [6; 7], так и потребность в формировании компетенций, связанных с навигацией по цифровым лексикографическим ресурсам, постепенно вытесняющим традиционные словари [8; 9; 10; 11]. Настоящее исследование направлено на анализ эффективности комплексного подхода, сочетающего использование автоматизированных систем с консультированием в онлайн-лексикографических источниках при работе с художественным текстом. В качестве гипотезы выдвигается предположение, что применение таких цифровых средств в совокупности позволяет будущим переводчикам успешно справляться даже с текстами повышенной сложности. Эмпирическую базу составляют фрагменты современной русской прозы, отобранные для выявления потенциала указанных инструментов в условиях, требующих точной передачи поэтических образов и культурно специфических реалий – элементов, с которыми автоматические системы, как показывают практика и исследования, сталкиваются с наибольшими затруднениями [12; 13; 14; 15].

Целью работы является выявление сильных и слабых сторон машинного перевода и оценка эффективности лексикографической поддержки при работе с переводом художественного текста. В задачи исследования входит: анализ типичных ошибок машинных переводчиков; выявление случаев, в которых лексикографическая проверка позволяет улучшить перевод и оценка достаточности электронных лексикографических ресурсов; оценка применимости цифровых инструментов в образовательном процессе.

В качестве текстов для анализа были отобраны фрагменты из произведений «Время: ночь» Людмилы Петрушевской, «Generation П» Виктора Пелевина и рассказов из сборника «Кысь» Татьяны Толстой. Все эти произведения представляют собой переходный этап от позднесоветской к постсоветской прозе, кроме того, в них встречается обилие языковых и культурных элементов. Для анализа были протестированы три популярных переводческих инструмента: Google Translate, Yandex Translate и ChatGPT-40. В качестве лексикографических источников использовались электронные версии толковых словарей русского языка (Ожегова, Ушакова), словарь RAЕ и другие лексикографические ресурсы, находящиеся в свободном доступе.

Первые фрагменты для анализа были заимствованы из произведения Людмилы Петрушевской «Время: ночь» [16]. Эти отрывки насыщены идиоматическими выражениями, сложной синтаксической структурой и характерной авторской лексикой, что делает их особенно трудными для машинного перевода. Ниже представлена сравнительная таблица автоматического перевода данного текста тремя системами: Yandex Translate, Google Translate и ChatGPT.

Таблица 1

Сравнительный анализ машинного перевода фрагментов повести
«Время: ночь» Л. Петрушевской тремя системами
(Yandex Translate, Google Translate, ChatGPT)

Оригинал	Google	Yandex	ChatGPT
(1) Он не ведаёт, что в гостях нельзя жадно кидаться к подзеркальнику и цапать все, вазочки, статуэтки, флакончики и особенно коробочки с бижутерией.	(1) Él no sabe que cuando visitas a alguien, no puedes correr con avidez al espejo y agarrar todo: jarrones, figuritas, botellas y, sobre todo, cajas de joyas.	(1) Él no sabe que en una visita no se puede tirar con avidez al espejo lateral y guardar todo, jarrones, estatuillas, frascos y especialmente cajas con joyas.	(1) No sabe que, cuando estás de visita, no se debe abalanzarse con avidez sobre el tocador ni agarrar todo lo que hay: jarroncitos, figuritas, frasquitos y, sobre todo, cajitas con bisutería.
(2) Владимир имеет нечто против Тимы, а Денис ему вообще надоел, как собака, зять, скажу я вам по секрету, явно на исходе, уже таёт, отсюда Оксанина ядовитость.	(2) Vladimir tiene algo contra Tima, y Denis le resulta completamente aburrido, como un perro, el yerno, te diré un secreto, está claramente en decadencia, ya se está derritiendo, de ahí el veneno de Oksana.	(2) Vladimir tiene algo en contra de tima, y Denis generalmente está cansado de él, como un perro, un yerno, te diré en secreto, obviamente se está agotando, ya se está derritiendo, de ahí la toxicidad de Oxanina.	(2) Vladímir tiene algo contra Tima, y Denis ya lo tiene harto, como un perro. El yerno, les digo en secreto, claramente está en las últimas, ya se está desvaneciendo, de ahí la ponzoñosa actitud de Oksana.

Анализ результатов выявил характерные особенности и недостатки машинного перевода на различных языковых уровнях.

На орфографическом уровне в системе Yandex Translate наблюдается некорректная передача собственного имени Оксана, представленная в виде *Oxanina*. Очевидно, система не распознаёт его как имя собственное. Это свидетельствует о низкой устойчивости механизма транслитерации имен.

На лексическом уровне существенные расхождения выявлены при передаче лексем с низкой частотностью и национально-культурной спецификой. Проанализируем несколько примеров. Так, слово *подзеркальник* в переводах Google и Yandex передается как *espejo* 'зеркало' и *espejo lateral* 'боковое

зеркало' соответственно, в то время как в версии ChatGPT используется *tocador* 'туалетный столик'. Обратимся к данным агрегатора онлайн-словарей Академик [17], где в разделе «Толкование» эта лемма представлена в 21 словарной статье. Согласно толковому словарю Ушакова, подзеркальник – это «столик, служащий подставкой для зеркала или прикрепленный к зеркалу-трюмо» [18]. Таким образом, вариант *tocador*, предложенный ChatGPT, является не только семантически точным, но и культурно адекватным.

Аналогичная тенденция наблюдается при передаче глагола *цанать*. В переводе, выполненном системой Yandex Translate, используется глагол *guardar* 'хранить', 'удерживать', 'уберегать', тогда как Google Translate и ChatGPT предлагают вариант *agarrar* 'хватать', 'схватить', 'цепляться'.

Глагол *цанать* обладает ярко выраженной экспрессивной окраской, указывающей на внезапность и резкость действия. Для верификации его значения обратимся к данным онлайн-словаря синонимов Sinonim.org [19], который ранжирует синонимы по частоте употребления и степени семантической близости. Согласно ресурсу, к числу синонимов глагола *цанать* относятся *брать*, *хватать*, *задевать*, *царапать* и другие, всего – 20 вариантов. Очевидно, что лексема *guardar*, предложенная Yandex, не только не совпадает по значению с исходным глаголом, но и искажает коммуникативное намерение высказывания. Следовательно, в данном случае можно говорить о лексической ошибке.

На морфологическом уровне заслуживает внимания передача уменьшительно-ласкательных форм существительных, таких как *вазочки*, *статуэтки*, *флакончики*. В переводах, выполненных системами Google Translate и Yandex Translate, данные формы частично нейтрализуются: уменьшительно-ласкательные суффиксы сохраняются лишь в слове *статуэтки* ('figuritas' / 'estatuillas'), тогда как *вазочки* и *флакончики* передаются как 'jarrones' и 'botellas' / 'frascos' – формы без экспрессивной маркированности. Напротив, перевод, предложенный ChatGPT, стремится сохранить морфологическую структуру оригинала, используя формы *jarroncitos*, *figuritas*, *frasquitos*, что позволяет передать эмоциональную окраску и авторскую интонацию, заложенную в исходном тексте.

На синтаксическом уровне представляет интерес анализ сложноподчиненного предложения из второго фрагмента и стратегий, применяемых различными системами перевода. Google Translate стремится сохранить исходную синтаксическую структуру, воспроизводя ее дословно и используя множество запятых. Yandex Translate придерживается аналогичного подхода: сохраняется линейное следование фраз. ChatGPT, напротив, принимает решение о синтаксической перестройке: исходное предложение делится на два более коротких, сохраняющих первоначальный смысл, но упрощающих восприятие текста на испанском языке. Однако, несмотря на то что передача содержания в данном случае остается точной, синтаксическое решение автора оригинального текста – сложное, ритмически напряженное высказывание – теряется, что приводит к изменению интонационной и стилистической организации текста.

На прагматическом уровне заслуживает особого внимания вопрос выбора формы обращения к адресату (*tú / usted / vosotros / ustedes*), так как он напрямую влияет на интерпретацию коммуникативной интенции и степень формальности высказывания. Так, при передаче конструкции *скажу я вам по секрету* Google Translate и Yandex Translate используют форму обращения на *tú* ('te diré un secreto', 'te diré en secreto'), что соответствует неформальному стилю общения. В то же время ChatGPT выбирает форму вежливого множественного числа – 'les digo en secreto', демонстрируя более формальный и дистанцированный регистр.

Кроме того, в конструкции *нельзя жадно кидаться к подзеркальнику* наблюдаются различия в стратегии передачи безличной модальности. Google Translate использует персонализированную форму 'no puedes correr con avidez', вводя субъекта (ты), чего нет в оригинале. Напротив, Yandex Translate и ChatGPT сохраняют безличный характер конструкции, применяя формы 'no se puede' и 'no se debe' соответственно. Последние решения представляются более адекватными с точки зрения прагматической нейтральности и соответствия исходной структуре.

Несмотря на то что в переводе данного фрагмента ChatGPT демонстрирует более точные и контекстуально обоснованные решения по сравнению с Google Translate и Yandex Translate, следует отметить, что и этот вариант не лишен недостатков и может быть улучшен с точки зрения стилистической и культурной адаптации. Так, эмоциональное выражение *надоел, как собака*, переданное в переводе ChatGPT дословно как 'como un perro', может быть воспринято испаноязычным читателем как неестественное или недостаточно идиоматичное. В испанском языке существуют более точные эквиваленты, передающие степень раздражения и экспрессии, например, *Estar hasta las narices / hasta el moño de alguien* – 'быть по горло сытым', буквально: 'до носа / до пучка волос' или *Me tiene frito / podrido* – 'он меня достал'. Использование подобных идиоматических конструкций позволило бы не только передать эмоциональный заряд оригинала, но и обеспечить функционально-стилистическое соответствие перевода, что особенно важно при работе с художественным текстом, насыщенным разговорной речью и экспрессивными оборотами.

Второй фрагмент анализа – отрывки из романа Виктора Пелевина «Generation „П“» [20], одного из представителей постсоветской прозы. Текст отличается насыщенностью философскими отсылками, игрой слов, неологизмами и элементами массовой культуры. Машинные переводчики испытывают затруднения при передаче многослойности смысла, иронии и интертекстуальных аллюзий, что позволяет выявить их ограничения в работе с художественным текстом высокой семантической плотности.

Таблица 2

Сравнительный анализ машинного перевода фрагментов романа
«Generation „П“» В. Пелевина тремя системами
(Yandex Translate, Google Translate, ChatGPT)

Оригинал	Google	Yandex	ChatGPT
(3) Когда-то в России и правда жило беспечальное юное поколение, которое улыбнулось лету, морю и солнцу – и выбрало «Пепси».	(3) Érase una vez en Rusia una generación joven despreocupada que sonreía al verano, al mar y al sol... y elegía Pepsi.	(3) Érase una vez en Rusia y realmente vivió una generación joven despreocupada que sonrió al verano, al mar y al sol, y eligió Pepsi.	(3) Hubo un tiempo en Rusia en el que realmente existía una generación joven y despreocupada, que sonrió al verano, al mar y al sol... y eligió Pepsi.
(4) И даже не в банальной взятке – хочется верить, что партийный бюрократ, от которого зависело заключение контракта, просто взял и полюбил эту темную пузырящуюся жидкость всеми порами своей разуверившейся в коммунизме души.	(4) Y ni siquiera un soborno banal : quiero creer que el burócrata del partido, de quien dependía la conclusión del contrato , simplemente fue y se enamoró de este líquido oscuro y burbujeante con cada poro de su alma, que había perdido la fe en el comunismo .	(4) Y ni siquiera en un soborno banal : uno quiere creer que el burócrata del partido, del cual dependía la conclusión del contrato , simplemente tomó y amó este oscuro líquido burbujeante por todos los poros de su alma, que había perdido la fe en el comunismo .	(4) Y ni siquiera se trata de un soborno banal – uno quiere creer que el burócrata del partido, de quien dependía la firma del contrato , simplemente se enamoró de ese líquido oscuro y burbujeante con cada poro de su alma desencantada del comunismo .

В переводах первого фрагмента наблюдаются существенные расхождения в переводческих решениях, принимаемых автоматизированными системами. Так, Google и Яндекс выбирают типичный сказочный зачин *Érase una vez* (аналог 'жили-были'). Учитывая, что данное предложение открывает текст всей книги, подобное вступление является оправданным, хотя и не обязательным. ChatGPT предлагает вариант *Hubo un tiempo en Rusia en el que...*, что дословно переводится как 'было время в России, когда...'. Такой вариант также допустим с точки зрения грамматики и семантики, однако обладает меньшей поэтичностью и стилистической маркированностью.

Наибольшие затруднения вызывает передача разговорных оборотов, характерных для русского языка. Так, выражение *и правда* Google полностью опускает, Яндекс передает как *y realmente*, что не является естественным выражением в испанском языке, тогда как ChatGPT сохраняет этот элемент,

предлагая дословный, но грамматически корректный вариант. Еще одним показательным примером является перевод оборота *взял и полюбил*. Google, Яндекс и ChatGPT переводят его как 'fue y se enamoró', 'tomó y amó' и 'se enamoró' соответственно. Несмотря на различия в выборе глаголов, Google и Яндекс в обоих случаях интерпретируют *взял* буквально – как отдельное целенаправленное действие. Между тем в русском языке конструкции типа *взял и* выполняют эмотивно-прагматическую функцию: согласно толкованию сервиса *Картаслов.ру* [21], сочетания с союзами *и*, *да*, *да и* и последующим глаголом выражают неожиданность или внезапность действия. Таким образом, только перевод, предложенный ChatGPT, адекватно отражает функционально-семантические особенности исходной конструкции, избегая кальки.

Дополнительные сложности вызваны выбором грамматического времени. Google в ряде случаев предпочитает pretérito imperfecto (*elegía, sonreía*), хотя контекст указывает на разовое завершённое действие, требующее употребления pretérito indefinido (*eligió, sonrió*). В этом отношении переводы Яндекса и ChatGPT выглядят более корректными с точки зрения грамматической нормы и соответствия коммуникативной установке оригинала.

Особый интерес представляет также передача безличной конструкции *хочется верить*. Google трансформирует ее в личную форму первого лица единственного числа – 'quiero creer', тем самым вводя субъективную модальность и авторское присутствие. Напротив, Яндекс и ChatGPT сохраняют безличную структуру, предлагая формулу 'uno quiere creer', что ближе к исходному русскому варианту как по значению, так и по стилистике. Таким образом, в рассматриваемом фрагменте именно Яндекс и ChatGPT демонстрируют наибольшую степень соответствия оригиналу как в передаче лексико-семантических элементов, так и в сохранении стилистической и прагматической целостности текста.

Третий фрагмент анализа – рассказы из сборника «Кысь» Татьяны Толстой [22], представительницы литературной элиты постсоветского периода. Ее стиль отличается высокой степенью языковой изобретательности, насыщен авторскими неологизмами, метафорическими конструкциями и элементами вымышленного постапокалиптического мира. Для корректного понимания текста требуется постоянная лексикографическая работа: сопоставление новых слов с контекстом, проверка их возможных прообразов и значений в современном русском языке. Такие условия предъявляют особые требования к переводчику и позволяют выявить различия в стратегиях автоматических систем при передаче нестандартных лексем и игровых элементов. Подобный материал становится особенно ценным для анализа, поскольку демонстрирует пределы адаптивности машинного перевода в условиях художественного текста с искусственно модифицированной языковой реальностью.

Таблица 3

Сравнительный анализ машинного перевода фрагментов рассказов
из сборника «Кысь» Т. Толстой тремя системами
(Yandex Translate, Google Translate, ChatGPT)

	Google	Yandex	ChatGPT
(5) Бенедикт натянул валенки, потопал ногами, чтобы ладно пришлось, проверил печную вьюшку, хлебные крошки смахнул на пол – для мышей, окно заткнул тряпицей, чтоб не выстудило, вышел на крыльцо и потянул носом морозный чистый воздух.	(5) Benedict se puso sus botas de fieltro, pateó el suelo para asegurarse de que estaban en buen estado, revisó el regulador de la estufa, barrió las migas de pan del suelo para los ratones, tapó la ventana con un trapo para protegerla del frío, salió al porche y olió el aire limpio y helado.	(5) Benedict tiró de las botas de fieltro, las hundió con los pies para que estuviera bien, revisó la bolsa del horno, deslizó las migas de pan en el piso, para ratones, tapó la ventana con un trapo para que no se apagara, salió al porche y tiró del aire limpio y helado con la nariz.	(5) Benedicto se calzó las valenkas, dio unos pisotones para que encajaran bien, revisó el tiro de la estufa, sacudió las migas de pan al suelo – para los ratones, tapó la rendija de la ventana con un trapo para que no entrara el frío, salió al porche y aspiró profundamente el aire limpio y helado.
(6) Бенедикт вздохнул: на работу пора; запахнул zipун, заложил дверь избы деревянным брусом и еще палкой подоткнул.	(6) Benedict suspiró: era hora de ir a trabajar; se abrochó el abrigo, bloqueó la puerta de la cabaña con una viga de madera y la sostuvo con un bastón.	(6) Benedict suspiró: es hora de trabajar; olió zipun, colocó la puerta de la cabaña con una barra de madera y abrió otro palo.	(6) Benedicto suspiró: era hora de ir al trabajo; se ajustó el zipún, atrancó la puerta de la isba con una viga de madera y la aseguró además con un palo.

В приведенном фрагменте наблюдается высокая насыщенность бытовыми деталями и национально окрашенной лексикой (валенки, печная вьюшка, хлебные крошки, тряпицей окно заткнул, потянул носом и т. д.).

Например, в случае со словом *валенки* Google и Яндекс предлагают перевод 'botas de fieltro', что соответствует определению данного вида обуви и может считаться приемлемым вариантом адаптированного перевода. Данный эквивалент подтверждается такими источниками, как статья из *The Fashionington Post*, где *valenki* определяются как 'botas de fieltro tradicionales rusas' [23]. Вместе с тем, ChatGPT транслитерирует слово как 'valenkas', что позволяет сохранить культурный колорит оригинала, однако такая форма представляет собой англоязычную кальку, не адаптированную к нормам

испанского языка, особенно в плане числа. Уместнее было бы использовать форму 'valenki', сохранив оригинальное написание без суффикса -s, с возможным кратким пояснением в сноске или скобках, чтобы облегчить понимание испаноязычному читателю без потери аутентичности.

Отдельного внимания заслуживает перевод словосочетания *печную вьюшку*, обозначающего, согласно Толковому словарю Ушакова, «покрышку, обыкновенно круглую чугунную, вкладываемую в печную трубу для прекращения тяги воздуха» [24]. Google предлагает вариант 'regulador de la estufa', что является технически точным и функциональным эквивалентом. ChatGPT использует выражение 'tiro de la estufa', которое также допустимо и звучит более естественно в разговорном регистре. Наиболее заметное отклонение наблюдается в переводе Яндексa – 'bolsa del horno', что представляет собой грубую ошибку: вместо заслонки (вьюшки) указывается предмет, не связанный с контекстом, а именно *пакет для запекания*. Этот пример наглядно демонстрирует важность семантической точности при переводе терминов, связанных с традиционным бытом и материальной культурой.

Интерес представляет и архаичное слово *zipun*, которое, согласно Толковому словарю живого великорусского языка В. И. Даля, означает «крестьянский кафтан без козыря» – простую верхнюю одежду, распространенную в дореволюционной России [25]. В выражении *запахнул zipun* Google предлагает перевод 'se abrochó el abrigo', что передает общее значение, но нейтрализует архаичный и культурно-исторический оттенок оригинала. ChatGPT использует вариант 'se ajustó el zipun', сохраняя форму и фонетику за счет транслитерации. Однако слово *zipun* отсутствует в испаноязычных словарях, что делает такой перевод затруднительным для понимания без дополнительных пояснений (например, в виде сноски или глоссария). Перевод Яндексa – 'olió zipun' – содержит фактическую ошибку: глагол *запахнул* был интерпретирован как *запахнул*, то есть в значении 'понюхал'. Это привело к выбору испанского *olió*, что указывает на сбой, вызванный омонимической интерференцией при распознавании ударения. Таким образом, данный пример иллюстрирует ограниченность автоматических переводчиков при работе с контекстуально и культурно маркированной лексикой.

В случае с переводом слова *изба* Google и Яндекс используют вариант 'la cabaña', в то время как ChatGPT предлагает 'isba'. Примечательно, что слово *isba* включено в словарную статью Diccionario de la lengua española Королевской академии [26], что делает его вполне уместным выбором в данном контексте. Хотя слово *cabaña* в целом передает основное значение – 'деревянное крестьянское жилище', наличие термина *isba* в номенклатуре одного из наиболее авторитетных испаноязычных словарей позволяет рассматривать его как предпочтительный вариант, особенно в контексте перевода, стремящегося сохранить культурно-историческую специфику оригинала.

На грамматическом уровне в переводе Яндексa наблюдается несогласование времен: фраза *Бenedикт вздохнул: на работу пора* передается как

'Benedict suspiró: es hora de trabajar', где глагол-сказуемое употреблен в прошедшем времени (*suspiró*), а следующая конструкция – в настоящем времени (*es hora*), что создает временной разрыв. Хотя грамматические ошибки как таковые отсутствуют у всех систем, на синтаксическом уровне выявляются различия в подходе к построению фразы. Google и Яндекс склонны к использованию кратких и раздельных конструкций, что способствует легкости восприятия, но одновременно снижает связность и стилистическую текучесть текста. В противоположность этому, ChatGPT демонстрирует более высокую степень владения синтаксисом испанского языка, активно выстраивая логические и стилистические связи между частями высказывания. Кроме того, в его переводах наблюдается использование сложных грамматических структур, включая сослагательное наклонение (*subjuntivo*), что приближает результат к литературной норме и делает текст более естественным для носителя языка.

Результаты

Анализ фрагментов из произведений Людмилы Петрушевской, Виктора Пелевина и Татьяны Толстой подтвердил, что эффективность машинного перевода (МП) напрямую зависит от степени языковой и культурной сложности текста. В случае нейтральной или официальной лексики все три системы (Google Translate, Yandex Translate и ChatGPT) демонстрируют высокую степень формального соответствия, однако при столкновении с идиоматикой, архаизмами, диалектизмами и стилистически маркированными структурами в художественных текстах выявляются существенные расхождения в качестве перевода.

На орфографическом уровне все системы, за редкими исключениями, справляются корректно. Ошибки возникают в основном при транслитерации имен и реалий (например, *Oxana* и *Oxxana*), что свидетельствует о необходимости настройки механизмов распознавания собственных имен и заимствованных слов.

Лексический уровень продемонстрировал наиболее яркие расхождения. В случае с реалиями советского и постсоветского быта (*валенки*, *печная вьюшка*, *zipun*) Google и Yandex склонны к нейтрализации или замене оригинальных лексем на более общие эквиваленты, в то время как ChatGPT, напротив, сохраняет или воссоздает их в виде транслитерации с адаптацией (*zipiún*, *valenkas*), что может быть интерпретировано как попытка культурной передачи. Однако такая стратегия требует дополнительной поддержки средствами сноски, синонимического пояснения или форматирования (курсив, сноски).

Морфологический и грамматический уровни в целом реализованы корректно во всех трех системах, однако наиболее стилистически гибкие решения наблюдаются в переводах, выполненных ChatGPT. Например, использование идиоматических конструкций, согласование времен и модальные

структуры отражают более высокий уровень контекстуальной осведомленности. Вместе с тем в ряде случаев наблюдаются семантические ошибки (например, *olió zírui* в Yandex), связанные с буквальным толкованием нестандартных или устаревших глаголов. Кроме того вызывает сложности система времен.

Синтаксически Google демонстрирует тенденцию к упрощению структуры: предложения разрываются, смысловые блоки теряют связность, что снижает ритм и экспрессию оригинала. ChatGPT сохраняет сложные синтаксические построения, в том числе авторский ритм и пунктуационные особенности (тире, вводные конструкции). Yandex, напротив, подчас нарушает логику действия, создавая синтаксически допустимые, но прагматически нелепые конструкции.

Прагматический уровень выявил существенные ограничения всех систем в передаче иронии, авторской позиции и культурной импликации. ChatGPT демонстрирует более высокую адаптивность к таким элементам, но при этом не всегда стабильно. Google и Yandex чаще стремятся к упрощению, что снижает глубину и стилистическую насыщенность текста. Таким образом, прагматическая адекватность перевода требует дополнительного вмешательства человека.

Обсуждение

Результаты анализа подтверждают, что даже наиболее продвинутые системы машинного перевода, включая нейросетевые модели с генеративной архитектурой, такие как ChatGPT, не обеспечивают гарантированной адекватности при переводе художественного текста. Особенно это проявляется в контексте работы с авторскими неологизмами, стилистически окрашенной лексикой, идиоматикой и иронией. Тем не менее, ChatGPT демонстрирует большую степень семантической гибкости и синтаксической вариативности по сравнению с более жестко алгоритмизованными решениями Google и Yandex.

При этом даже наиболее «успешный» машинный перевод требует редактирования, особенно если речь идет об образовательной или профессиональной практике. Лексикографическая поддержка – в виде словарей, контекстных платформ и справочных ресурсов – оказывается критически важной для корректной интерпретации и стилистической передачи оригинала. В условиях подготовки переводчиков это подчеркивает необходимость интеграции критического анализа машинного перевода в учебные программы.

Таким образом, в условиях цифровой трансформации лингвистического образования машинный перевод представляет собой полезный инструмент, способный существенно облегчить начальные этапы перевода, особенно в учебной среде. Однако он не может заменить навыки интерпретации, понимания культурного контекста и стилистического выбора, которыми обладает профессиональный переводчик. Оптимальным подходом представляется

сочетание МП и лексикографической проверки в рамках системной подготовки студентов к работе с художественными текстами. Особое внимание должно быть уделено развитию навыков критического анализа результатов автоматического перевода, в том числе через сравнительный и дидактический подход.

ЛИТЕРАТУРА

1. Дзюба А. В. Новые технологии и перевод // Практики и интерпретации: журнал филологических, образовательных и культурных исследований. 2016. № 1 (3). С. 166–176.
2. Гарбовский Н. К. «Цифровой перевод». Современные реалии и прогнозы // Русский язык и культура в зеркале перевода: IX Международная научная конференция (г. Салоники, Греция, 10–14 апреля 2019 г.): материалы конференции. М. : Изд-во Московского университета, 2019. С. 65–72.
3. Кольцова Д. А. Формирование переводческих компетенций в цифровую эпоху // Вестник Московского университета. Сер. 22. Теория перевода. 2022. № 2. С. 114–124.
4. Гао Цзяи. Структура переводческой деятельности в эпоху цифровизации // Litera. 2022. № 10. С. 72–86.
5. Безобразова Т. А. Влияние цифровых технологий на обучение переводчиков: от традиционного к цифровому подходу // Вестник науки. 2024. № 3 (11 (80)). С. 769–779.
6. Котенко В. В. Перспективы развития нейронного машинного перевода в контексте концепции открытого образования // Ученые записки университета им. П. Ф. Лесгафта. 2020. № 4 (182). С. 225–231.
7. Раренко М. Б. Машинный перевод: от перевода «по правилам» к нейронному переводу // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 6, Языкознание: реферативный журнал. 2021. № 3. С. 70–79.
8. Ридецкая Ю. С. Неография и интернет-лексикография // Фортунатовские чтения в Карелии: сб. докл. междунар. науч. конф., Петрозаводск, 10–12 сент. 2018 г. / под ред. Н. В. Патроевой. Петрозаводск: ПетрГУ, 2018. Ч. 2. С. 102–104.
9. Лексикография цифровой эпохи: сб. материалов междунар. симпозиума (24–25 сент. 2021 г.) / отв. ред. Е. А. Юрина, С. С. Земичева. Томск: Изд-во Томского гос. ун-та, 2021. 420 с.
10. Воронцова И. А. Интернет-словари, сетевые решения и инструменты автоматизации как компоненты электронного рабочего места переводчика // Научно-исследовательская деятельность в классическом университете: традиции и инновации: материалы междунар. науч.-практ. фестиваля, Иваново, 19–29 апр. 2021 г. Иваново: Иван. гос. ун-т, 2021. С. 412–416.
11. Рябцева Н. К. Современная лексикография, дополненная лингвистическая реальность и проблемы терминологии // Лингвистика и методика преподавания иностранных языков. 2022. № 1 (16). С. 75–110.

12. Мироненко А. Б., Омельченко П. А. Проблемы машинного перевода // Актуальные проблемы авиации и космонавтики. 2016. Т. 2 (12). С. 1193–1195.
13. Лосева С. А. Перевод и искусственный интеллект: проблемы и пути развития // Межкультурная коммуникация и профессионально ориентированное обучение иностранным языкам: материалы XVI Междунар. науч. конф., Минск, 27 окт. 2022 г. Минск: БГУ, 2022. С. 162–168.
14. Козьякова М. В., Корнеева И. В. Творческий потенциал ChatGPT в современной практике перевода корейских поэтических текстов на примере жанра сиджо // Язык как структура и дискурсивная практика. 2023. № 2 (2). С. 18–26.
15. Джунковский А. В. Искусственный интеллект как новая форма САТ. Путь к идеальной автоматизации перевода или новый вызов? // Речевые технологии. 2024. № 1. С. 14–17.
16. Петрушевская Л. С. Время: ночь / Л. С. Петрушевская. – М. : АСТ, 2018.
17. Академик: агрегатор онлайн-словарей [Электронный ресурс]. – URL: <https://academic.ru/> (дата обращения: 04.06.2025).
18. Подзеркальник // Ушаков Д. Н. Толковый словарь русского языка [Электронный ресурс]. – URL: <https://dic.academic.ru/dic.nsf/ushakov/944474> (дата обращения: 04.06.2025).
19. Sinonim.org: Словарь синонимов русского языка [Электронный ресурс]. – URL: <https://sinonim.org/> (дата обращения: 04.06.2025).
20. Пелевин В. О. Generation «П» / В. О. Пелевин. – М. : Эксмо, 2003.
21. Значение слова «взять» // Картаслов.ру [Электронный ресурс]. – URL: <https://kartaslov.ru/значение-слова/взять> (дата обращения: 29.06.2025).
22. Толстая Т. Н. Кысь: рассказы / Т. Н. Толстая. – М. : Захаров, 2001.
23. Valenki // The Fashionington Post [Электронный ресурс]. – 2024. – 2 июля. – URL: <https://fashioningtonpost.com/es/valenki/> (дата обращения: 29.06.2025).
24. Вьюшка // Ушаков Д. Н. Толковый словарь русского языка [Электронный ресурс]. – URL: <https://dic.academic.ru/dic.nsf/ushakov/1341615> (дата обращения: 04.06.2025).
25. Даль В. И. Зипун // Толковый словарь живого великорусского языка [Электронный ресурс]. – URL: <https://dic.academic.ru/dic.nsf/enc2p/245362> (дата обращения: 29.06.2025).
26. Real Academia Española. Isba // Diccionario de la lengua española [Электронный ресурс]. – URL: <https://dle.rae.es/isba> (дата обращения: 22.06.2024).

УДК 372.881.111.1

Разумова Анна Викторовна

старший преподаватель кафедры
лексикологии и стилистики
английского языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Hanna Razumava

Senior Lecturer at the Department
of English Lexicology and Stylistics
Belarusian State University
of Foreign Languages
Minsk, Belarus
Razumova_anna@mail.ru

ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ОБУЧЕНИИ СТУДЕНТОВ ПЕРЕВОДУ

Данная статья посвящена проблеме использования программ и приложений искусственного интеллекта для преподавания дисциплин переводческого цикла. В статье приводятся примеры таких приложений, которые дают возможность предредактирования выбранного материала, непосредственно его перевода, а также постредактирования. В статье приводятся примеры заданий, которые могут быть использованы преподавателем для более эффективной работы студентов в аудитории и самостоятельно. Для развития переводческих навыков у студента предлагается широкий набор упражнений, который включает индивидуальную, парную и групповую работу студентов. Обучающимся каждый раз предлагается выполнить перевод либо самостоятельно, либо проанализировать и сравнить переводы, выполненные различными программами ИИ. Автор подчеркивает значимость использования новых технологий для более эффективной работы преподавателя и студентов, однако, указывает, что все приложения и программы ИИ лишь помощь в процессе преподавания, а не замена традиционным методическим подходам. Автор статьи дает ссылки на современные программы и приложения ИИ, которые могли бы быть использованы в курсе преподавания дисциплин переводческого цикла.

Ключевые слова: искусственный интеллект; обучение переводу; контекст; автоматизация; машинный перевод.

POSSIBILITIES OF USING ARTIFICIAL INTELLIGENCE IN TEACHING STUDENTS TRANSLATION

This article is devoted to the problem of using artificial intelligence programs and applications for teaching translation-related subjects. The article provides the examples of such applications that enable pre-editing of the selected material, its proper translation and post-editing. The article contains the examples of the tasks that can be used by the teacher for more effective work of students in the classroom and independently. To develop the students' translation skills, a wide range of exercises is offered, which include individual, paired, and group work of students. Each time, students are asked to perform the translation either independently or to analyze and compare translations made by different AI programs. The author emphasizes the importance of using new technologies for more effective work of the teacher and students and at the same time points out that all AI applications and programs are only an aid in the teaching process, and not a replacement for traditional methodological approaches. The author of the article provides links to modern AI programs and applications that could be used in the course of teaching translation-related subjects.

Key words: artificial intelligence; teaching translation; context; automation; machine translation.

Современные технологии, в частности искусственный интеллект (ИИ), активно внедряются в различные сферы жизни общества: медицину, производство, искусство, а также образование. В процессе преподавания дисциплин лингвистического характера, все чаще в последние годы можно заметить, как и студенты и преподаватели обращаются к приложениям ИИ для создания, выполнения, проверки и анализа заданий. Дисциплины переводческой направленности не являются исключением. Однако, обучение переводу – сложный процесс, требующий не только глубокого понимания языков, но и культурных особенностей, контекста и специализированной терминологии. Традиционные методы преподавания перевода сталкиваются с рядом ограничений, таких как недостаток персонализации, высокая нагрузка на преподавателей и сложность оперативной проверки работ [1]. Таким образом, искусственный интеллект имеет возможность предложить инновационные решения, способные значительно улучшить процесс обучения будущих учителей, преподавателей, переводчиков. В данной статье мы предлагаем рассмотреть ключевые возможности применения ИИ в обучении переводу, включая автоматизированный анализ текстов, адаптивное обучение, использование нейросетей и оценку качества перевода студентов.

Студенты факультета английского языка изучают две дисциплины переводческого цикла: в 8 семестре они знакомятся с переводом с английского на русский на дисциплине «Введение в переводоведение» и продолжают перевод с русского на английский в 9 семестре с дисциплиной «Основы перевода».

В системе подготовки специалистов с высшим образованием в соответствии с учебными планами направления специальностям 1-21 06 01-01 «Современные иностранные языки (преподавание)» и 1-02 03 08 «Иностранный язык (с указанием языка)» учебные дисциплины «Введение в переводоведение» и «Основы перевода» являются дисциплинами компонента учреждения образования цикла специальных дисциплин и составной частью модуля «Перевод».

Целью изучения данных учебных дисциплин является формирование и развитие основных навыков и умений одностороннего перевода (с английского языка на русский или наоборот), необходимых для решения задач в сфере профессиональной деятельности.

В ходе преподавания данных дисциплин, основным является решение следующих задач:

- 1) ознакомить с переводом как особым видом коммуникативной деятельности и основными понятиями теории перевода;
- 2) раскрыть взаимосвязь теории перевода с другими дисциплинами лингвистического цикла и активизировать применение знаний, полученных в языковой практике;
- 3) сформировать навыки, необходимые для достижения эквивалентности в переводе при помощи лексических и грамматических переводческих трансформаций;

4) развить навыки отбора языковых средств для адекватной передачи содержания переводимого текста с учетом его прагматической заданности и жанрово-стилевой принадлежности.

В системе подготовки специалистов с высшим образованием в соответствии с учебным планом специальности 1-02 03 08 «Иностранный язык (с указанием языка)» учебная дисциплина «Основы перевода» является дисциплиной компонента учреждения высшего образования. В общей системе профессиональной подготовки преподавателей английского языка содержание учебной дисциплины «Основы перевода» базируется на знании учебных дисциплин «Функциональная грамматика», «Лексикология», «Стилистика». В результате изучения данных учебных дисциплин студенты должны знать:

- основные термины и понятия науки о переводе;
- лексико-грамматические трансформации (приемы), применяемые в практике перевода;
- основные грамматические категории в родном и иностранном языках и особенности их передачи в письменном переводе;
- способы достижения эквивалентности в письменном переводе.

После изучения данных курсов студенты должны уметь работать со словарями различного типа (двуязычными, толковыми, тезаурусными), в том числе электронными, передавать при переводе имена собственные, реалии, безэквивалентную лексику, термины, фразеологические единицы, стилистические средства, выполнять письменный перевод публицистического, научно-популярного или художественного текста с английского языка на русский. Следующие сформированные умения включают: осуществление при переводе необходимой прагматической перестройки текста, анализ, редактирование и комментарий текста перевода [2; 3].

Очевиден тот факт, что при выполнении заданий на перевод из предлагаемых учебников, студенты активно используют автоматические переводчики, а также программы ИИ. Для того, чтобы максимально задействовать именно потенциал студента, вызвать его интерес к самостоятельной, а не автоматической работе на занятиях предлагается использовать комбинированные способы работы с текстом, а именно систему студент плюс ИИ.

Сперва следует рассмотреть автоматизированный анализ и предпереводческую подготовку текста оригинала. Как нам видится, одним из важных этапов обучения переводу является работа с исходным текстом: его анализ, выявление ключевых терминов, стилистических особенностей и потенциальных сложностей. ИИ может значительно упростить этот процесс за счет:

- лингвистического анализа (определение грамматических структур, тональности, сложности текста);
- терминологического поиска (автоматическое выделение и объяснение специализированных терминов) [4];
- сравнения с параллельными корпусами текстов (поиск аналогичных переводов в базах данных). На данном этапе студенты используют анализ

уже существующих переводов, выделяют сильные и слабые стороны своего перевода. Здесь уместно использовать такие ресурсы как *Sketchengine* [5] и *Laurenceanthony* [6].

Существуют и инструменты на основе NLP (Natural Language Processing), которые могут автоматически размечать текст, выделяя идиомы, многозначные слова и сложные синтаксические конструкции, что помогает студентам лучше понять исходный материал перед началом перевода. Одним из таких инструментов является сайт *Voyant-tools* [7].

Современные системы машинного перевода (Google Translate, Chatgpt, Deepseek) очень часто используются студентами особенно во время выполнения домашних заданий. Безусловно, эти программы достигли высокого уровня качества, однако их прямое использование без критического осмысления может лишь навредить обучению. Тем не менее, ИИ-переводчики могут быть полезны в учебном процессе, если применять их правильно.

После выполнения перевода, ИИ может быть использован как инструмент сравнения – студенты могут анализировать машинный перевод, находить ошибки и предлагать свои более точные варианты, здесь есть возможность как индивидуальной работы студентов, так и сотрудничества в группе. Использование ИИ для постредактирования машинного перевода может быть источником развития навыков критического мышления и улучшать понимание нюансов языка [8]. Прекрасные ресурсы для грамматической, стилистической проверки предоставляет приложение Grammarly, используя которое, студенты задействуют и свои компетенции в области грамматики, стилистики, лексикологии и других смежных дисциплин [9].

Преподаватели могут давать студентам задания, в которых имеется возможность сравнивать свой перевод с ИИ-версией, аргументируя выбор тех или иных решений, подкрепляя данный вид деятельности проектной работой, используя игровые методы, предлагая студентам быть критиками, анализировать, обосновывать выбор того или иного варианта перевода.

Особенностью преподавания дисциплин переводческого цикла на факультете английского языка является то, что каждый студент обладает разным уровнем владения языком и скоростью усвоения материала. ИИ позволяет создавать персонализированные программы обучения за счет использования индивидуального подхода при обучении переводу. Преподаватель может с минимальными затратами времени и ресурсов генерировать индивидуальные задания для студентов, размещать их на платформе Moodle. В зависимости от прогресса студента ИИ может подбирать тексты соответствующей сложности и тематики [10]. Такие платформы, как *Duolingo* или *Chatgpt* уже используют элементы адаптивного обучения, и подобные технологии могут быть доработаны для профессиональной переводческой подготовки.

Следующим немаловажным этапом сотрудничества в аудитории является оценка качества перевода и обратная связь. Проверка переводов – одна из самых трудоемких задач для преподавателя. ИИ может автоматизировать этот процесс с помощью:

– сравнения с эталонными переводами, где алгоритмы оценивают близость студенческого перевода к "идеальному" варианту. Здесь себя отлично зарекомендовала платформа *Smartcat* [11];

– генерации развернутых комментариев – современные ИИ (*GPT-4*) способны не только находить ошибки, но и объяснять, как их исправить.

Это не заменяет работу преподавателя, но сокращает время проверки и позволяет уделять больше внимания индивидуальным консультациям, подготовке к занятиям, самосовершенствованию преподавателя.

Помимо чисто технической стороны перевода, ИИ может помочь в развитии творческого подхода у студентов, что крайне важно при подготовке будущих специалистов в сфере образования, педагогов. ИИ может имитировать заказчика, давая реальные технические задания и задавая условия к ним. Здесь мы видим и дальнейшую возможность студентов использовать полученные навыки при написании научных исследований, анализируя большие массивы текстов, поиск закономерностей в переводах.

Таким образом, как пример можно предложить следующие задания в ходе обучения студентов переводу.

1. *Загрузите текст в ИИ-анализатор (предложить на выбор).*
2. *Выделите ключевые термины, многозначные слова и идиомы.*
3. *Составьте глоссарий перед переводом.*

Данные задания ставят цель научиться анализировать текст перед переводом.

Для сравнения стилистических особенностей текстов перевода и оригинала можно предложить следующие виды деятельности.

1. *Сравните два текста на одном языке (например, научный и публицистический стиль).*
2. *Определите частотность слов, тональность, сложность предложений.*
3. *Переведите тексты с учетом их стилистики.*

Упражнения на критический разбор ИИ-перевода могут выглядеть следующим образом.

1. *Переведите текст вручную.*
2. *Сравните с машинным переводом.*
3. *Найдите ошибки и неточности, объясните, почему ваш вариант лучше.*

Задания на постредактирование машинного перевода.

1. *Возьмите машинный перевод технического текста.*
2. *Исправьте ошибки, адаптируйте под естественный язык.*
3. *Оцените, сколько времени сэкономил ИИ.*

Современные технологии позволяют преподавателю создавать и групповые упражнения с ИИ. В Google Docs студенты вместе редактируют один текст с ИИ-поддержкой. Здесь вновь предлагается использовать ресурс Moodle, который дает возможность размещать подобного рода задания. После чего студенты могут обсуждать спорные моменты и принимать коллективное решение.

Как мы уже упоминали, искусственный интеллект открывает новые горизонты в обучении переводу, предлагая инструменты для автоматизации рутинных задач, персонализации обучения и развития профессиональных навыков. Однако его использование должно быть осмысленным. ИИ – не замена преподавателю, а мощный вспомогательный инструмент. В будущем интеграция ИИ в переводческое образование будет только усиливаться, и ключевой задачей педагогов станет обучение студентов критическому взаимодействию с технологиями, чтобы они могли не просто полагаться на машинный перевод, а использовать его для повышения собственной квалификации. ИИ-инструменты предоставляют мощную поддержку в обучении переводу, начиная от анализа текстов и заканчивая автоматизированной проверкой. Однако их следует использовать осознанно, сочетая с традиционными методами обучения. Преподаватели могут интегрировать эти программы в учебный процесс, чтобы сделать его более интерактивным и эффективным.

Таким образом, для максимальной пользы рекомендуется сравнивать машинный и человеческий перевод, использовать ИИ для тренировки постредактирования, применять адаптивные платформы для персонализации обучения.

ИИ позволяет разнообразить обучение переводу, предлагая интерактивные упражнения (разбор ошибок, постредактирование), персонализацию (адаптивные тренажеры, карточки), креативные задачи (генерация вариантов, имитация заказчика).

Соответственно, основными задачами для преподавателя видятся следующие: комбинирование ИИ с традиционными методами, акцентирование внимание на критическом анализе, а не слепом доверии к машинному переводу. Что очень важно при большой нагрузке – использование ИИ для рутинных задач, чтобы освободить время для творческой работы.

ЛИТЕРАТУРА

1. Bowker L. Machine translation // Taylor & Francis eBooks. 2021. Гл. 7. URL: <https://www.taylorfrancis.com/chapters/oa-mono/10.4324/9781003217718-7/machine-translation-lynne-bowker>. – DOI: 10.4324/9781003217718-7/ (дата обращения: 12.07.2025).
2. Ломовая А. В. Введение в переводоведение : пособие. Минск : Изд-во МГЛУ, 2020. 180 с.
3. Электронный учебно-методический комплекс по учебной дисциплине «Основы перевода» для специальности 1-02 03 08 «Иностранный язык (английский)» [Электронный ресурс] / Минский гос. лингвист. ун-т ; сост.: А. В. Ломовая, А. В. Драневич. Минск, 2023. URL: <http://elib.mslu.by/handle/edoc/12277/> (дата обращения: 11.07.2025).
4. Cabré M. T. Terminology : theory, methods and applications // Terminology. Amsterdam : John Benjamins Publishing, 1999. Vol. 5, № 1. P. 1–12. ISSN 0929-9971.

5. Sketch Engine : корпусный менеджер для работы с текстами. URL: <https://www.sketchengine.eu> (дата обращения: 10.07.2025).
6. Anthony L. AntLab: Laurence Anthony's Website. URL: <https://www.laurenceanthony.net/> (дата обращения: 10.07.2025).
7. Voyant Tools: визуальная среда для анализа текстов. URL: <https://voyant-tools.org> (дата обращения: 11.07.2025).
8. Moorkens J. Machine translation and the translation profession // The Journal of Specialised Translation (JoSTrans). 2018. № 30 (July). P. 132–148. ISSN 1740-357X. URL: https://www.jostrans.org/issue30/art_moorkens.pdf (дата обращения: 12.07.2025).
9. Grammarly: Free AI Writing Assistance. URL: <https://www.grammarly.com> (дата обращения: 11.07.2025).
10. Luckin R. Exploring the future of learning and the relationship between human intelligence and AI: An interview with Professor Rose Luckin [Electronic resource] // Journal of Applied Learning and Teaching. 2025. URL: <https://journals.sfu.ca/jat/index.php/jalt/article/view/1659/> (date of access: 11.07.2025).
11. Smartcat CAT-инструмент: автоматизированный перевод с ИИ // Smartcat. URL: <https://ru.smartcat.com/cat-tool/> (дата обращения: 11.07.2025).

УДК 81

Салахиева Анастасия Юрьевна
магистрант
Санкт-Петербургский
Политехнический университет Петра
Великого
г. Санкт-Петербург, Россия

Anastasia Salakhieva
Graduate student
Peter the Great St. Petersburg
Polytechnic University
Saint Petersburg, Russia
n.salahieva@mail.ru

ИСПОЛЬЗОВАНИЕ СИСТЕМ НЕЙРОННОГО МАШИННОГО ПЕРЕВОДА ДЛЯ ПЕРЕВОДА КОМИЧЕСКОГО (на примере анимационного телесериала «Inside Job»)

В статье рассматривается использование систем нейронного машинного перевода (НМП) для перевода комического на примере анимационного телесериала «Inside Job». Актуальность исследования обусловлена ростом популярности аудиовизуальных произведений и необходимости сохранения их культурных и стилистических особенностей при переводе. Особое внимание уделяется особенностям дубляжа, а также специфике передачи комического в переводе, который основан на национально-специфических элементах и использовании различных стилистических средств. В качестве материала исследования выбран американский сериал «Inside Job», для которого проведен сравнительный анализ перевода с помощью систем ChatGPT, GigaChat и DeepL. В статье было изучено развитие технологий машинного перевода и перспективы его использования в аудиовизуальном переводе. В результате исследования были выявлены особенности использования НМП при переводе комического, сделаны выводы о перспективах применения НМП в аудиовизуальном переводе.

Ключевые слова: нейронный машинный перевод; аудиовизуальный перевод; дубляж; комическое; стилистические средства.

THE USE OF NEURAL MACHINE TRANSLATION FOR TRANSLATING COMIC ELEMENTS (the Animated Series “Inside Job”)

This article examines the use of neural machine translation (NMT) systems for translating comedic content, using the animated television series "Inside Job" as a case study. The relevance of the research is driven by the growing popularity of audiovisual works and the need to preserve their cultural and stylistic features in the process of translation. Special attention is given to the specifics of dubbing, as well as the challenges of conveying humor in translation, which relies on culturally specific elements and various stylistic devices. The American animated series "Inside Job" was selected as the material for analysis, with a comparative study conducted on translations produced by ChatGPT, GigaChat, and DeepL systems. The article explores the development of machine translation technologies and their prospects for application in audiovisual translation. As a result of the research, particular features of NMT use in translating humor were identified, and conclusions were drawn regarding the future potential of NMT in audiovisual translation.

Key words: neural machine translation; audiovisual translation; dubbing; comic; stylistic means.

С середины XX века существует и активно развивается идея применения дополнительных технологий для осуществления перевода без участия человека. За последнее время был достигнут значительный прогресс в разработке и внедрении технологий машинного и автоматизированного перевода, однако в то же время стал чаще подниматься вопрос о качестве такого перевода. Благодаря развитию цифровых технологий и их внедрению в переводческую деятельность перед переводчиками появляются новые задачи. Один из таких вызовов – применение генеративного искусственного интеллекта и систем машинного перевода для создания эквивалентных текстов различных жанров с исходного языка на язык перевода.

Машинный перевод зародился в 1954 году. Тогда была представлена первая система перевода вычислительной техники – «Джорджтаунский эксперимент». С тех пор системы машинного перевода активно совершенствовались и изменялись. Первоначально использовался буквальный, пословный перевод, который часто приводил к искажению смысла текста. Дальнейшее развитие привело к созданию более сложной системы перевода, при которой учитывалась синтаксическая структура и правила грамматики двух языков. Тем не менее, данный прорыв не решил проблемы сохранения семантики текста оригинала.

На сегодняшний день существует несколько основных типов систем машинного перевода: основанный на правилах, основанный на статистике, и нейронный машинный перевод. Первый тип систем схож с системами второго поколения. Среди них можно выделить два подвида: трансферные и интерлингвистические. Они связаны общим принципом работы: связь структуры входного и выходного предложения. К концу прошлого века стал активно использоваться статистический подход машинного перевода. Основным принцип данного типа перевода основывается на обучении машины на основе баз параллельных текстов, в которых одинаковая информация на разных языках изучалась на уровне слов и предложений [2, с. 2]. Однако за последние пару лет стала применяться новая технология – нейронный машинный перевод (далее – НМП). НМП основан на нейронных сетях. Эта система обучена на базах данных параллельных текстов и представляет собой предсказание последовательностей слов. Этот процесс аналогичен переводу человека. Сначала НМП «читает» исходное предложение, затем, основываясь на своем понимании предложения, производит новое предложение. По сравнению с предыдущими методами НМП не нуждается в специальном обучении правилам языков, разработанных человеком, она сама находит закономерности при обучении на базах данных [14, р. 190]. Благодаря этим преимуществам НМП в настоящее время является доминирующим методом осуществления машинного перевода. НМП уже используют такие программы, как Google Переводчик, Яндекс Переводчик, DeepL и другие [7, с. 71].

Несмотря на то, что применение НМП стало прорывом в сфере технологий и представляет широкие возможности для применения и развития, она обладает рядом значительных недостатков.

Во-первых, в НМП предложение является основным элементом декодирования, поэтому могут возникать ошибки при переводе, так как значение слова в предложении может быть определено только в контексте. Помимо этого, при переводе документа необходимо сохранять единообразие перевода терминов, в то время как выполнение перевода предложения за предложением независимо зачастую не позволяет достичь такого результата. Также наблюдаются сложности в декодировании идиоматических выражений, что приводит к искажению смысла высказывания.

Во-вторых, нейронный перевод доступен не для всех языковых пар. Крупные базы данных параллельных текстов существуют только в нескольких языковых парах, поэтому проблема отсутствия корпусов для параллельного обучения распространена в большинстве языков и областей, что создает ограничения для применения НМП. Более того, разные языки имеют разную структуру и правила, что делает перевод между языками сложной задачей [5, с. 45].

Также НМП часто совершает ошибки и подвержен «галлюцинациям», поэтому экспертам необходимо тщательно проверять полученный результат.

Расширение сферы применения НМП не могло не затронуть аудиовизуальный перевод. На данный момент все активнее изучается вопрос применения НМП для осуществления голосового перевода (дубляжа) и субтитрования. Отдельно стоит отметить, что многие социальные сети с аудиовизуальным материалом, такие как «Instagram» (принадлежит компании Meta, признанной экстремистской и запрещенной на территории РФ), «ВКонтакте» и другие, внедряют функцию распознавания речи и генерирования субтитров на английском и русском языке соответственно. Ряд других веб-страниц используют искусственный интеллект для перевода аудиовизуальных материалов на другие языки, среди них крупнейшая видео-платформа «Youtube», которая предоставляет автоматическое субтитрование и голосовой перевод, «Sreep» и другие. Среди отечественных веб-браузеров, обладающих подобным функционалом, можно выделить «Яндекс Браузер» [4, с. 574–575]. Тем не менее, как отмечают исследователи, существует острая проблема повышения качества такого перевода для адекватной передачи основного содержания сообщения.

Далее рассмотрим особенности перевода аудиовизуального материала. Перевод – это сложный многосторонний процесс, который включает в себя передачу смысла, нюансов, структуры и стиля текста или устного высказывания из одного языка на другой. Он играет важную роль в обмене информацией, и содействии межкультурной коммуникации. Переводчик должен учитывать контекст, цель и целевую аудиторию перевода, а также сохранять точность и стилистическое соответствие.

Особенно актуальной областью для исследования становится аудиовизуальный перевод. В связи с увеличением объема контента, производимым теле- и киностудиями, все большую популярность приобретает киноиндустрия, производимый киностудиями аудиовизуальный контент и его изучение. В частности, полнометражные кинопроизведения и сериалы.

Воспользуемся определением аудиовизуального перевода, которое предложил российский переводчик А. В. Козуляев: «Аудиовизуальный перевод – это создание нового полисемантического единства на языке-реципиенте на основе единства, существовавшего на исходном языке, причем таким образом, чтобы новое полисемантическое единство стало элементом культуры языка-реципиента и не было ему чуждо» [6, с. 13].

Испанский лингвист Патрик Сабальбеаскоа считает, что аудиовизуальный текст должен содержать смесь вербальных, невербальных, визуальных и аудиоэлементов, обладающих равной значимостью [15, р. 29].

Говоря об особенностях аудиовизуального перевода, А. В. Козуляев подчеркивает, что переводчик работает с четырьмя каналами данных, среди которых: 1) визуальным невербальным образным рядом; 2) невербальным шумомузыкальным аудиорядом; 3) вербальным аудиорядом (диалоги героев); 4) вербальным видеорядом (надписи на экране, субтитры) [6, с. 5].

Так, аудиовизуальные произведения и их перевод изучаются как зарубежными, так и российскими исследователями. Особенности поликодовых текстов, АВП и его виды освещены в работах таких отечественных ученых, как А. В. Козуляев, Е. Д. Маленова, а также зарубежных исследователей: Х. Диаз-Синтаса [12], П. Сабальбеаскоа [15] и других.

К видам аудиовизуального перевода относятся озвучивание и субтитрование. К озвучиванию, в свою очередь, можно отнести дубляж и закадровый перевод (voiceover). В данной работе мы сосредоточимся на переводе под дубляж.

Перевод под дубляж – это такой вид перевода, при котором оригинальная акустическая вербальная дорожка (то есть, слова) «изымается» и заменяется вербальной дорожкой на языке перевода, таким образом, что реципиент воспринимает получившееся аудиовизуальное произведение как продукт своей собственной культуры. При этом перевод под дубляж предполагает соблюдение трех видов синхронизации: линсинк (lipsync), то есть синхронность звука с артикуляцией, синхронность звука и движения, жестов и изохронность (синхронность звука и просодических особенностей, таких как паузы, длительность высказываний и т. д.). Другими словами, артикуляция переводимого высказывания, его просодия, а также кинетические знаки (жесты), производимые говорящими и необходимые для понимания сюжета, должны быть эквивалентно переданы на языке перевода [6, с. 19].

Дублирование является способом сделать фильм менее «чужим» и иностранным, так как актеры говорят на том же языке, что и зритель, а иностранные высказывания приводятся в соответствие с внутренними культурными

нормами. Даже если в фильме использован иностранный диалект, он должен быть адаптирован к приближенным эквивалентам, принятым в целевой культуре.

Часто анимационные фильмы характеризуются упрощенной, единообразной прорисовкой артикуляции персонажей, когда персонаж либо открывает, либо закрывает рот без движения губ. Однако некоторые мультипликационные фильмы и сериалы (в том числе и анализируемый в данной статье мультсериал) отличаются более детальной прорисовкой артикуляции, при которой персонажи артикулируют такие звуки как [о], [у], [и], [ы], [в], [ф]. В какой-то степени такая «неполная» артикуляция несколько облегчает работу переводчика. Тем не менее, такое произношение необходимо учитывать, чтобы в итоге три вида синхронизации были соблюдены.

Комическое является важной частью коммуникации и может создаваться с помощью специальных вербальных и невербальных средств. Поэтому перевод комического требует использования особых приемов при перекодировании на другой язык. Комическое – явление общечеловеческое, в котором проявляются как общие, так и частные законы юмора, и в то же время это – явление, определяемое особенностями национального характера, культурными традициями и социальным устройством. При передаче комического эффекта с одного языка на другой переводчик сталкивается с трудностями экстралингвистического, культурологического и собственно лингвистического характера, требующими при переводе комплексного использования различных типов преобразований.

Также не менее важно, что термин «комическое» часто используется как синоним понятий «смешное» и «юмор». Некоторые исследователи (А. Д. Кошелев и т. д.) разделяют данные понятия. Во-первых, это связано с более широкой сферой употребления понятия «смех», а также возможностью употребления термина «юмор» в узком смысле для обозначения одного из видов комического [8, с. 45].

В целом в лингвистике под комическим понимают лингвистический феномен, основанный на противоречии и заключающийся в реализации эффекта обманутого ожидания, коммуникативная цель которого рассмешить адресата.

За рубежом работы по исследованию категории комического принадлежат Роду Мартину, Виктору Раскину, Сальваторе Аттардо и др. Особенности перевода аудиовизуальных произведений с комическими элементами рассматривали О. В. Мишина, Е. Д. Юрченко и др.

Понимание комического, его правильное восприятие и интерпретация во многом зависит от коммуникативных ожиданий адресата и его фоновых знаний. При этом категория комического имеет национально-культурную специфику, то есть представления о смешном у представителей разных народов и эпох будут разными. Эта особенность представляет особую трудность при переводе, однако она должна быть учтена для эквивалентной передачи коммуникативного эффекта.

Как уже было сказано ранее, существуют вербальные и невербальные средства создания комического. Первые строятся на использовании языковых единиц (например, жаргонизмов, профессионализмов, специальных терминов, фразеологизмов) и средств художественной выразительности. Невербальные средства основываются на использовании визуального и аудиального компонентов (гримасы, жесты, шум, грохот и пр.). В данной статье мы сосредоточимся на вербальных, то есть языковых средствах, с помощью которых достигается комический эффект.

Анализ работ Н. А. Абросимовой, М. Р. Желтухиной, Е. Е. Жук и И. В. Цикушевой [9, с. 101] позволяет заключить, что языковыми средствами создания комического являются стилистические приемы и выразительные средства. Кроме того, их можно классифицировать по уровням языка, опираясь на типологию, предложенную И. Р. Гальпериным. Так как уровни или группы, на которые перечисленные авторы разделяют все средства, различаются в их классификациях (например, И. В. Цикушева выделяет фонетический, лексико-семантический, морфологический, стилистико-грамматический уровни; Е. Е. Жук – лексические, синтаксические и звуковые средства, и т. д.) [13, с. 3], имеет смысл составить свою классификацию на основе предложенных.

Итак, в данном исследовании будет использоваться следующая классификация языковых средств создания комического эффекта:

- 1) фонетические (ономатопея, аллитерация, ассонанс, рифма);
- 2) лексические (метафора, метонимия, каламбур, гипербола, зевгма, оксюморон, аллюзии к культурным реалиям, парадокс, сарказм, сравнение);
- 3) грамматические (морфологические и синтаксические) (окказионализмы).

Как уже было сказано ранее, тесная связь комического с культурными особенностями одного народа, которые безусловно отражаются и в языке, создает трудности при переводе комического, то есть трудности для его передачи средствами другого языка.

Л. С. Бархударов утверждал, что при переводе комического важно не только передать смысл, понятный для зрителя, но и сохранить юмористический элемент [1, с. 142]. В своих работах П. А. Лоу отмечает, что каждая шутка должна быть переведена, но переведена так, чтобы вызвала тот же эффект, который закладывал автор текста на языке оригинала [13, р. 62]. Для этого переводчики используют различные стилистические средства и трансформации.

Таким образом, можно заключить, что адекватный перевод единиц с комическим эффектом – это такой перевод, который вызывает у адресата перевода аналогичную с адресатом оригинала реакцию.

Цель исследования: проанализировать использование систем нейронного машинного перевода для перевода комического (на примере анимационного телесериала «Inside Job»).

Объект исследования: системы нейронного машинного перевода.

Предмет исследования: перевод комического с помощью систем нейронного машинного перевода на примере анимационного телесериала «Inside Job».

Задачи исследования: изучить развитие технологий нейронного машинного перевода и их применение в сфере аудиовизуального перевода, рассмотреть понятие комического и определить элементы комического в анимационном телесериале «Inside Job», проанализировать качество перевода комического, выполненного системами НМП ChatGPT, GigaChat, DeepL.

Материалом для исследования стал сериал «Корпорация заговор» (англ. Inside Job). Это американский комедийный и научно-фантастический анимационный сериал для взрослой аудитории (18+), спродюсированный Сион Такэути для сервиса Netflix в 2021 году. Официальный перевод под дубляж был выполнен под руководством Бориса Новикова. Сериал был озвучен студией КупиГолос в 2021 году [10]. Сюжет разворачивается вокруг девушки-инженера и ее коллег, работающих в американской организации Cognito Inc. Эта компания принадлежит Теневому правительству и призвана скрывать от людей подтверждения теорий заговора. Данный сериал представляет интерес для исследования, в том числе и с точки зрения перевода, так как комический эффект в сериале часто достигается посредством использования в речи персонажей американских реалий, которые, как известно, могут вызывать трудности при переводе, особенно аудиовизуальном [11].

Для проведения анализа нами были отобраны примеры передачи комического на разных языковых уровнях, в соответствии с разработанной нами классификацией. Было проведено сравнение переводов, выполненных системами НМП ChatGPT, GigaChat и DeepL, а также официального перевода студии КупиГолос. Выбор данных систем был обоснован широким распространением данных систем. ChatGPT – продукт компании Open AI, выпущенный в 2022 году. GigaChat – русскоязычная нейросеть от компании «Сбербанк», представленная в 2023 году. DeepL – платформа языкового искусственного интеллекта, которая обучается на собственных языковых данных компании. Более того, он пользуется большой популярностью у профессиональных переводчиков. Стоит отметить, что ChatGPT и GigaChat позволяют задать текстовых запрос с объяснением целей и контекста перевода, в то время как данная функция отсутствует в переводчике DeepL. Данный факт представляет дополнительный интерес для исследования и сравнения трех систем. При анализе всех примеров была сохранена пунктуация и орфография НМП.

Перейдем к анализу примеров из анимационного телесериала «Inside Job». На лексическом уровне самыми частотными приемами создания комического являются аллюзии на реалии, создание каламбуров и парадоксов. Рассмотрим несколько примеров:

Glenn Dolphman: I pledge the Fifth and the Second.

В данном эпизоде одного из героев сериала, Гленна Дельфмана, вызвали в суд в качестве свидетеля для дачи показаний. Однако он не хочет отвечать

на заданные вопросы, и поэтому говорит клишированную фразу, используемую в суде: *I pledge the Fifth. 'Взываю к Пятой поправке'*. Данная поправка к Конституции США гарантирует возможность не свидетельствовать против себя и хранить молчание. Однако юмористический элемент добавляется, когда главный герой продолжает свое высказывание и говорит, что хочет воспользоваться также Второй поправкой США и после этого достает оружие. Вторая поправка к Конституции гарантирует право гражданина на хранение и ношение оружия. Жители США знают значение основных поправок и с легкостью поймут почему Гленн достал оружие после произнесенной фразы, однако для российского зрителя будет непонятна взаимосвязь между двумя действиями, если только он не знаком со значениями этих поправок. Таким образом, комический элемент создается за счет аллюзии на американскую реалию.

Я пользуюсь Пятой поправкой и Второй поправкой. (ChatGPT)

Я обещаю Пятую и Вторую. (DeepL)

Заявляю Пятую поправку... и заодно Вторую! (GigaChat)

Итак, при переводе данной реплики Chat GPT и GigaChat добавили слово “поправка” для конкретизации смысла для русскоговорящего зрителя, что помогает сохранить ключевую отсылку на реалию США. Тем не менее, в переводе ChatGPT присутствует необоснованный лексический повтор. В варианте, предложенном GigaChat и DeepL, неверно подобран эквивалент слова *pledge*. В то же время в переводе DeepL отсутствует лексическое добавление в виде слова *поправка*, в связи с чем теряется отсылка на Конституцию США. Таким образом, потребуется постредактирование переводов, чтобы зритель смог верно понять предложение и отсылку, благодаря которой создается комический эффект.

Гленн Дельфман: Взываю к Пятой поправке и ко Второй. (КупиГолос)

В дублированном тексте данная фраза передается с помощью буквального перевода и добавления слова «поправка». Таким образом, отсылка на реалию сохраняется.

Рассмотрим еще одно частое средство достижения комического эффекта – использование парадокса, как в следующем примере:

Glenn Dolphman: I had to reheat the Cold War.

В данной сцене сериала герои рассказывают о том, какие методы они использовали, чтобы отвлечь внимание от неудачной миссии и скрыть ее от населения. Один из персонажей, Гленн, высказывает вышеприведенную фразу. Алогизм и наличие противоречия, свойственные парадоксу, основаны, во-первых, на противоположной семантике слов *reheat* и *cold* и во-вторых, на использовании прилагательного в словосочетании *cold war* одновременно и в прямом, и в переносном значении вместе с глаголом *reheat*, тоже употребленном в прямом значении. Как следствие возникает резкое расхождение с привычным, устоявшимся понятием о холодной войне, как об идиоматическом описании политической ситуации.

Мне пришлось повторно разогреть Холодную войну. (ChatGPT)

Пришлось заново разогреть «холодную войну». (DeepL)

Приходилось разогревать Холодную войну заново. (GigaChat)

Во всех переводах, выполненных НМП, были допущены орфографические или пунктуационные ошибки при переводе выражения «холодная война». Корректным вариантом является написание выражения с маленькой буквы и без кавычек. Стоит отметить, что в переводе DeepL и GigaChat был использован глагол несовершенного вида *разогревать* и глагол прошедшего времени *приходилось*, которые показывают, что действие выполнялось несколько раз, что нарушает логику высказывания.

Гленн Дельфман: Вот я холодную войну разогрел. (КупиГолос)

В официальном переводе парадокс сохранился, что позволило передать комический эффект высказывания.

Говоря о приемах создания комического на грамматическом уровне языка, было выявлено, что персонажи сериала нередко прибегают к созданию окказионализмов, заменяя часть слова на другое слово на основе созвучности. Рассмотрим следующий пример.

Brett: Yeah, not gonna lie, even for you, that's Reaganier than usual.

В данном случае окказионализм основан на грамматике английского языка, а именно на образовании сравнительной степени наречия с помощью суффикса *-er*, а также на метафорическом переносе значения странности, необычности на слово Рейган из-за аналогичного поведения главной героини. То есть, имя Рейган приобретает добавочное значение странности, становится синонимично наречиям «странно, необычно». Используя слово Рейган в качестве наречия и добавляя к нему суффикс сравнительной степени, Брэтт, один из героев, называет поведение главной героини еще более странным, чем обычно.

Далее рассмотрим переводы, выполненные НМП:

Да, не буду врать, даже для тебя это более Рейгановское, чем обычно. (ChatGPT)

Да, не буду врать, даже для тебя это Рейгана больше, чем обычно. (DeepL)

«Честно скажу, даже для тебя это было чересчур „рейганисто“, чем обычно». (GigaChat)

Как мы можем заметить, при переводе используется попытка сохранить окказионализм, добавляя к имени Рейган суффиксы. В первом переводе, выполненном ChatGPT, создается прилагательное от имени героини, которое в сочетании со словом *более* нарушает грамматику русского языка. Более того, в данном варианте присутствует семантическая ошибка, так как непонятно к чему относится данное прилагательное без существительного. В переводе, выполненном DeepL, окказионализм *Reaganier* был переведен неверно, в связи с чем создается отсылка на фамилию президента США Рональда Рейгана, что нарушает изначальный смысл предложения. К тому же, в данном варианте перевода нарушены синтаксис и грамматика русского языка, из-за чего само высказывание теряет смысл. В варианте перевода GigaChat для образования окказионализма к имени героини добавляется суф-

фикс -ист, с окончанием наречий -о. Однако данный окказионализм звучит «чуждо» в переводе. К тому же словосочетание *чересчур «рейганисто»* нарушает нормы грамматики русского языка.

Рассмотрим официальный перевод.

Брэтт: Врать не буду, даже для тебя это как-то необычно (КупиГолос)

Ввиду различий грамматического строя английского и русского языка, переводчикам не удалось использовать аналогичный прием для создания окказионализма. В переводе данная шутка объясняется, то есть она передается с помощью экспликации и понижения образности. В результате комический эффект утрачивается, но смысл высказывания остается.

Проанализируем еще несколько окказионализмов.

Reagan: Wow, free moon-iversity? Moon-iversal healthcare? A living moon-imum wage?

В данном эпизоде сериала главные герои отправляются на Луну, где встречают людей, живущей на ней. Базз Олдрин, один из лунных жителей, объясняет, что у жителей Луны есть одно правило: добавлять к любому высказыванию слова, связанные с космосом и Луной. Таким образом, на протяжении всей серии разные герои создают окказионализмы, добавляя слово *moon* 'Луна'. Как мы видим, в данном примере, слово *moon* было использовано вместо первых слогов слов *university*, *universal* и *minimum*.

Проанализируем переводы НМП.

Рейган: Вау, бесплатный луно-университет? Луно-здравоохранение? Живой луно-минимальный заработок? (ChatGPT)

Рейган: Ух ты, бесплатный лунный университет? Всеобщее лунное здравоохранение? Живая луна-минимальная зарплата? (DeepL)

«Ух ты, бесплатный лунверситет? Лунверсальная медпомощь? Животрепещущая луннимая ставка зарплаты?» (GigaChat)

Все системы НМП предприняли попытку сохранить окказионализмы при переводе на русский язык. ChatGPT образовал окказионализмы со словами *университет*, *здравоохранение* и *минимальный*, добавив к ним морфему -луно-, однако была сохранена орфография языка оригинала. В то же время в переводе DeepL и GigaChat отсутствует единообразие при переводе и создании окказионализмов. В двух случаях перевода слово *moon* было переведено как прилагательное *лунный*. Также был создан окказионализм с добавлением слова *Луна*. В примере GigaChat были созданы два окказионализма с добавлением морфем *лунь* и *лун*, однако слово *moon* в последнем окказионализме было переведено как *луннимый*. Более того, НМП сохраняют часть орфографии языка оригинала, в связи с чем дефис ставится после морфемы -лун-. Системы не смогли распознать выражение *living minimum wage* 'прожиточный минимум', в котором первый слог *min* был заменен на *moon*. В связи с чем слово *living* было переведено буквально, таким образом была нарушена семантика исходного выражения.

Рейган: Ого, бесплатный луниверситет, луниверсальная больница, лунимальная оплата труда? (КупиГолос)

В примере официального перевода переводчикам удалось сохранить прием и создать окказионализмы на русском языке со словами той же семантики, что и в оригинале.

Рассмотрим еще один пример.

Locals: Have a great after-moon.

Buzz Aldrin: Don't moon-tion it.

Во втором примере слово *moon* использовано вместо первого слога слова *mention* и второго слога слова *afternoon*, так как в них имеются похожие или одинаковые звуки. Проанализируем переводы, выполненные НМП.

Местные: Хорошего луно-времени.

Базз Олдрин: Нелуно-упоминай. (ChatGPT)

Местные жители: Хорошего послеполуденного отдыха.

Базз Олдрин: Не надо луны. (DeerL)

Жители Луны: Хорошего вам послеполумя!

Базз Олдрин: И не надо лунтуировать. (GigaChat)

В каждом из примеров системы НМП попытались сохранить окказионализм, добавив морфему -лун- к словам. Тем не менее, ChatGPT сохранил буквальный перевод фразы *Don't mention it*, добавив к ней корень “луно-”, из-за чего нарушается прагматика сообщения, так как смысл фразы оказывается передан неверно. Перевод реплики Базза Олдрина, выполненный DeerL, также не является эквивалентным, так как смысл фразы оказывается искажен, нарушается прагматический эффект оригинала и грамматика языка перевода. При переводе высказывания GigaChat создал два окказионализма с морфемой -лу- и -лун-, однако их значения непонятны в данном контексте и искажают прагматику исходного сообщения оригинала.

Местные: Хорошего лунодня!

Базз Олдрин: Луните на здоровье! (КупиГолос)

В примере официального перевода переводчикам удалось сохранить прием и создать окказионализмы на русском языке со словами той же семантики, что и в оригинале.

На фонетическом уровне в исследуемом языковом материале мультсериала наиболее частотным стилистическим приемом создания комического эффекта является рифма, например:

Мус: This fungi is ready for fun, guys!

В данной серии герои сериала отправляются на ежегодную галацеремонию теневого правительства. Перед входом в зал проведения мероприятия один из героев говорит вышеприведенную фразу, чтобы продемонстрировать свое воодушевление. Комичность достигается за счет рифмы, в основе которой лежит игра слов *fungi* и *fun, guys*, сходство их фонетической формы

[ˈfʌŋg(a)ɪ] и [fʌn] [gaɪz], а также созвучие окончаний. Стоит также отметить, что персонаж, который говорит данную реплику, является грибом. Таким образом, данный элемент необходимо сохранить при переводе.

Рассмотрим переводы НМП.

Этот гриб уже готов к веселью, ребята! (ChatGPT)

Этот гриб готов к веселью, ребята! (DeepL)

«Грибочки наши танцуют веселый танец, братцы!» (GigaChat)

Мы можем заметить, что ChatGPT и DeepL буквально перевели данную фразу, в связи с чем рифма не сохранилась. В то же время GigaChat попытался сохранить рифму, был использован прием целостного преобразования, однако полученная реплика не обладает необходимым прагматическим эффектом.

Майк: Ваш друг грибок / обещает, что все будет ок. (КупиГолос)

Стоит отметить, что в официальном переводе комический эффект в высказывании персонажа был сохранен и воссоздан также посредством создания рифмы. Несмотря на то, что изменяются лексические компоненты, прагматика сохраняется.

Таким образом, системы машинного перевода развиваются с начала прошлого века, и на данный момент нейронный машинный перевод все активнее используется в разных областях перевода. Передача комического элемента при переводе на иностранный язык остается одной из сложных задач для переводчика. Тем не менее, системам нейронного машинного перевода не удастся совершить перевод комического с сохранением ключевых комических элементов. Более того, в результате перевода нарушается грамматика русского языка и прагматика сообщения. При сравнении трех систем, использующих НМП, было выявлено, что переводчик DeepL хуже справляется с задачей распознавания смысла сообщения из-за отсутствия возможности составить текстовый запрос с объяснением значения переводного элемента. В то же время несмотря на то, что для получения эквивалентного варианта перевода в ChatGPT и GigaChat были отправлены текстовые запросы с объяснением значений выражений и контекста, они не смогли декодировать значения выражений и сохранить элемент комического. Более того, НМП не доступен анализ дополнительных аудиовизуальных каналов, которые являются важными при выполнении перевода и передачи комического элемента. Таким образом, нужно с особым вниманием подходить к использованию НМП для перевода комического в аудиовизуальных произведениях. Экспертам придется прибегнуть к постредактированию перевода для передачи нужного прагматического эффекта.

ЛИТЕРАТУРА

1. Бархударов Л. С. Язык и перевод (Вопросы общей и частной теории перевода) / Л. С. Бархударов. – М., 1975. – 240 с.

2. Дроздова К. А. Машинный перевод: история, классификация, методы // Вестник Омского государственного педагогического университета. Гуманитарные исследования. 2015. № 3 (7). URL: <https://cyberleninka.ru/article/n/mashinnyy-perevod-istoriya-klassifikatsiya-metody>.
3. Жук Е. Е. Особенности изучения языковых средств комического и их перевода (на примере произведений О. Генри и П. Г. Вудхауса) // Вестник Адыгейского государственного университета. Серия 2: Филология и искусствоведение. 2012. № 2. URL: <https://cyberleninka.ru/article/n/osobennosti-izucheniya-yazykovykh-sredstv-komicheskogo-i-ih-perevoda-na-primere-proizvedeniy-o-genri-i-p-g-vudhausa>.
4. Калинин А. Ю. Автоматический голосовой перевод в социальных медиа: анализ воспринимаемого качества перевода тревел-влогов платформы YouTube // Медиа лингвистика, 2023. 10(4). С. 568–597. URL: <https://doi.org/10.21638/spbu22.2023.408>.
5. Касьянов В. К., Федулова В. В. Основные проблемы нейронного машинного перевода // Успехи в химии и химической технологии. 2021. № 11 (246). URL: <https://cyberleninka.ru/article/n/osnovnye-problemy-neyronnogo-mashinnogo-perevoda>.
6. Козуляев А. В. Обучение динамически эквивалентному переводу аудиовизуальных произведений: опыт разработки и освоения инновационных методик в рамках Школы аудиовизуального перевода // Вестник ПНИПУ. Проблемы языкознания и педагогики. 2015. № 3 (13). URL: <https://cyberleninka.ru/article/n/obuchenie-dinamicheski-ekvivalentnomu-perevodu-audiovizualnyh-proizvedeniy-opyt-razrabotki-i-osvoeniya-innovatsionnyh-metodik-v>.
7. Искусственный интеллект в технологиях машинного перевода / Колин К. К., Хорошилов А. Д., Никитин Ю. В., Пшеничный С. И., Хорошилов А. А. // Социальные новации и социальные науки. 2021. № 2. С. 64–80. URL: <https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-tehnologiyah-mashinnogo-perevoda>.
8. Кошелев А. Д. О структуре комического (анекдот, каламбур, шарж, пародия, шутка, комическая история) // Логический анализ языка / отв. ред.: Н. Д. Арутюнова. – М.: Изд-во «Индрик», 2007. С. 325.
9. Цикушева И. В. Лингвистические средства создания комического эффекта в сказках // Вестник АГУ. Серия «Филология и искусствоведение». Майкоп, 2008. – Выпуск 10. – С.: 100–103.
10. КупиГолос : сайт. Москва. Обновляется в течение суток. URL: <https://kupigolos.ru/kto-ozvuchivaet/multiserial-korporaciya-zagovor>.
11. Кинопоиск : сайт. Москва. Обновляется в течение суток. URL: https://www.kinopoisk.ru/film/1263653/cast/who_is/translator/.
12. Díaz Cintas J., Introduction – Audiovisual Translation: An Overview of its Potential // New Trends in Audiovisual Translation – Bristol; Buffalo; Toronto. P.: 1–18.

13. Low P.A. Translating jokes and puns // Perspectives: Studies in Translatology. 2011. Vol.:19(1). P.59–70. DOI: 10.1080/0907676X.2010.493219.
14. Orel Kos S., Introduction of Machine Translation into Audiovisual Translation Teaching // ELOPE: English Language Overseas Perspectives and Enquiries. 2024. Vol.:21(1). P.185–208. URL: <https://doi.org/10.4312/elope.21.1.185-208>.
15. Zabalbeascoa P., Translation in constrained communication and entertainment // Díaz Cintas J., Matamala A., Neves J., (Eds.) New insights into audiovisual translation and media accessibility: Media for all 2 – Amsterdam; New York. P.25–40. DOI:10.1163/9789042031814_004.

УДК 81'255.2:004.8

Чернова Ирина Валериевна

кандидат филологических наук,
доцент кафедры теории и практики
перевода

Кубанский государственный
университет

г. Краснодар, Россия

Irina Chernova

PhD in Philology,
Associate Professor of the Department
of Translation

Kuban State University

Krasnodar, Russia

chernova.iv.elt@mail.ru

Динаева Карина Муратовна

студентка 5 курса кафедры теории
и практики перевода

Кубанский государственный
университет

г. Краснодар, Россия

Karina Dinaeva

Senior student of the Department
of Translation

Kuban State University

Krasnodar, Russia

dinaeva.karina@ya.ru

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ПЕРЕВОДЕ ХУДОЖЕСТВЕННЫХ ТЕКСТОВ

Данная статья посвящена изучению места искусственного интеллекта при осуществлении художественного перевода и выработке рекомендаций по совершенствованию обучения нейронных машинных переводчиков, призванных улучшить качество перевода художественных текстов. Актуальность настоящего исследования обусловлена несколькими факторами: важностью определения роли искусственного интеллекта и человека в переводе художественных текстов, необходимостью оценки уровней эквивалентности и адекватности переводов, выполняемых с использованием систем нейронного машинного перевода, расширением классификации ошибок нейросетевого перевода, разработанной Евгением Матусовым, а также требованиями современных государственных образовательных стандартов, предписывающих подготовку высокопрофессиональных конкурентоспособных специалистов в области лингвистики и перевода, умеющих использовать современные информационно-коммуникационные технологии в профессиональной переводческой деятельности.

Ключевые слова: искусственный интеллект; нейронный машинный перевод; художественный перевод; художественный текст; эквивалентность; адекватность.

ARTIFICIAL INTELLIGENCE IN LITERARY TEXT TRANSLATION

The article is devoted to the study and evaluation of the role of artificial intelligence in literary translation, as well as the development of recommendations for improving the training of neural machine translation systems to enhance the quality of literary text translations. The relevance of the study is determined by several factors: the importance of defining the roles of artificial intelligence and humans in translation of literary texts, the need to assess the levels of equivalence and adequacy in translations produced by neural machine translation systems, the expansion of the classification of neural translation errors developed by Evgeny Matusov, and

the requirements of modern state educational standards, which mandate the training of highly qualified and competitive specialists in linguistics and translation who are capable of using modern information and communication technologies in professional translation activities.

Key words: artificial intelligence; neural machine translation; literary translation; literary text; equivalence; adequacy.

Новизна темы исследования заключается в ее актуальной междисциплинарности и недостаточной изученности в аспекте оценки возможности и степени успешности использования искусственного интеллекта и, в частности, нейронных машинных переводчиков при переводе художественных текстов.

Цель исследования состоит в оценке качества работы нейронных машинных переводчиков на двух уровнях – уровне адекватности и эквивалентности перевода через анализ ошибок, допускаемых системами при переводе художественных текстов, а также разработка рекомендаций для переводчиков, занимающихся художественным переводом.

Теоретическую базу настоящего исследования составили труды в области художественного перевода, компьютерной лингвистики, искусственного интеллекта, машинного обучения Т. А. Казаковой [1], О. И. Коваль [2], С. А. Лосевой [3], Е. А. Морозкиной [4], Ш. Ю. Мусоевой [5], L. Hinkel [6].

Материалом исследования послужили тексты произведений “The Hitchhiker’s Guide To The Galaxy” Дугласа Адамса [7] и “The Girl Who Circumnavigated Fairyland On The Ship Of Her Own Making” Кэтрин Валенте [8] на английском языке и их переводы, выполненные В. Бакановым [9], В. Беленковичем [10] и системами нейронного машинного перевода (далее НМП) «Яндекс Переводчик», Google Translate и DeepL.

В рамках исследования для каждой системы НМП был проведен анализ 100 текстовых фрагментов с подсчетом допущенных ошибок. Объем выборки составил 156 предложений, отобранных по таким критериям, как наличие в предложениях общенаучной лексики и специальных терминов, авторских неологизмов, образных лексических единиц, характеризующих жанры научной фантастики и фэнтези, к которым относятся анализируемые произведения, и представляющим потенциальную сложность для перевода средствами искусственного интеллекта. При проведении настоящего исследования мы опирались на специфику жанра научно-фантастической литературы, особенности художественного текста и художественного перевода, характеристику категорий оценки качества перевода (эквивалентность и адекватность), и использовали классификацию ошибок нейросетевого перевода, предложенную Евгением Матусовым [11], который выделяет 10 категорий ошибок:

- M1: серьезные смысловые ошибки (severe meaning error);
- M2: незначительные смысловые ошибки (minor meaning error);
- U: ошибки неизвестного слова или сегментации (unknown word or segmentation error);
- C: ошибки в переводе терминов (consistency / term translation error);
- P: ошибки в определении местоимений (pronoun resolution error);

- L: ошибки в переводе идиом (locution error);
- O, I, R: ошибки в опущении, дополнении, повторении; (omission, insertion, repetition errors);
- S1: серьезные синтаксические ошибки (severe syntax error);
- S2: незначительные синтаксические ошибки (minor syntax error);
- T: ошибки тона / регистра (tone / register error).

Рассмотрим примеры перевода фрагментов произведений, выполненных нейронными машинными переводчиками. Первый пример посвящен тому, как были переданы системами названия книг, которые Дуглас Адамс упоминает в произведении, противопоставляя их известность популярности книги “The Hitchhiker’s Guide to the Galaxy”:

“Not only is it a wholly remarkable book, it is also a highly successful one – more popular than ‘*the Celestial Home Care Omnibus*’, better selling than ‘*Fifty More Things to do in Zero Gravity*’, and more controversial than Oolon Colluphid’s trilogy of philosophical blockbusters ‘*Where God Went Wrong*’, ‘*Some More of God’s Greatest Mistakes*’ and ‘*Who is this God Person Anyway*’” – «Книга пользуется огромным спросом, оставив далеко позади «Межзвездный сборник советов по домоводству»; раскупается гораздо быстрее, чем «Еще 53 способа скоротать время в невесомости», и вызывает еще больше споров и толков, чем философская трилогия «Где Господь сбился с пути», «Основные ошибки Господа Бога» и «Да кто он в конце концов такой, этот Господь Бог?!» (перевод В. Баканова) [9].

Так, в названии книги *The Celestial Home Care Omnibus* словосочетание *home care* Яндекс перевел как «на дому», допустив незначительную синтаксическую ошибку, так как корректнее было бы использовать «по до́му» в значении ухода за домом. Этот вариант использовали переводчики Google и DeepL. Также, Яндекс опустил в переводе “celestial”, переведя предложение как «Сборник по уходу на дому», а DeepL транслитерировал “omnibus”, переведя название книги как «Небесный омнибус по уходу за домом», где слово «омнибус» явно незнакомо русскоязычному читателю. В переводе названия книги *Who is this God Person Anyway* важно было передать пренебрежительный оттенок словосочетания *god person*. Яндекс и Google использовали в переводе «Божественный человек», что придает фразе излишнюю сакральность и относится к ошибке регистра T, а также не соотносится с предыдущим употреблением слова «Бог». При этом Google использует «вообще», что помогает сохранить определенную фамильярность, связанную с отношением автора к названному опусу.

В произведении Кэтрин Валенте, главная героиня, девочка Сентябрь, попадает в волшебное измерение, где встречает различных существ. Сестры-ведьмы, с которыми она познакомилась в лесу, рассказывают девочке о мифическом происхождении своего мужа, который у сестер один на двоих. Он был *wairwolf* – волком, в полнолуние обращающимся в человека. Слово было придумано автором произведения для создания сказочного таинст-

венного образа персонажа. Все системы перевели его как «оборотень», от англ. *werewolf*, что нарушает задумку автора и искажает игру слов. В то время как переводчик справился с задачей и передал понятие через «зверовольф», тем самым, так же, как и автор произведения, создав новое слово, что современным нейронным системам сделать не под силу.

Другим переводческим препятствием для нейросетей стал перевод имен собственных и дальнейшее сохранение единообразия перевода имен при последующем упоминании персонажей. Все переводчики продемонстрировали тенденцию к буквальной интерпретации имен собственных. Так, например, имя “Hello” было переведено как «Хелло» (Яндекс, Google), «Алло» (DeepL), что превращает имена в обычные реплики, лишая их художественной многозначности. Наиболее проблемным оказался перевод имени “Manythanks”: НМП-системы либо неудачно транскрибировали его («Мани-танк» – DeepL, «МэнниСпасибо» – Google), либо переводили дословно («большое спасибо» – Яндекс), полностью игнорируя контекстную игру слов. Так как произведенные нейросетями замены имен вводят читателя в заблуждение и требуют обращения к оригиналу, то их можно отнести к категории серьезных смысловых ошибок. К сожалению, при переводе имен собственных и авторских неологизмов системы нейронного машинного перевода продемонстрировали отсутствие способности к творческому преобразованию и низкий уровень эквивалентности.

Распределение ошибок, допускаемых НМП при переводе художественных текстов, по категориям позволило выявить самые частотные. Так, в процессе анализа переводов были обнаружены незначительные смысловые ошибки, ошибки в тоне/регистре, что доказывает тот факт, что НМП испытывает трудности в сохранении нюансов замысла автора, даже при формальной точности перевода. Общей тенденцией для всех НМП стали и несерьезные стилистические ошибки (S2).

Одной из наиболее распространенных ошибок, допускаемых НМП была ошибка, связанная с распознаванием отношения местоимений. Так, грубую ошибку допустила система DeepL при переводе следующих предложений:

“At the top of the podium loomed an enormous *gargoyle*, *its* face a mass of bronze and black rock, wagging stone eyebrows and a stern metal jaw. *Its* lolling eyes burned red flames” – «На вершине подиума возвышалась огромная *горгулья* с лицом из бронзы и черного камня, виляющими каменными бровями и суровой металлической челюстью. *Его* запавшие глаза горели красным пламенем» (перевод В. Баканова).

В русском языке существительное *горгулья* женского рода, соответственно местоимение, с которого начинается второе предложение следует перевести местоимением женского рода (*ее*). Но как мы видим, если при переводе первого предложения системе удалось обойти употребление местоимения «горгулья с лицом», то при переводе второго предложения у DeepL не получилось связать местоимение *its* с существительным *горгулья* и перевести его местоимением женского рода, а не мужского. В данном случае, мы

также можем наблюдать такое явление как предвзятость, когда система в большинстве случаев предпочитает перевести местоимение *it/its* местоимением мужского рода, а не женского. Данную проблему выделяют многие исследователи, занимающиеся изучением этических вопросов, связанных с искусственным интеллектом [12], [13].

Сравнение систем нейронного машинного перевода привело нас к следующим выводам:

- DeepL лучше сохраняет тон, склонен к добавлениям, и чаще ошибается в идиомах (L) и терминах (C). Например, он неверно истолковал выражение *“to be on a home stretch”*, как «иметь все шансы на успех». В то же время, другие переводчики применили эквивалентное выражение «выйти на финишную прямую».

- Google Translate минимизирует грубые смысловые ошибки (M1), но лидирует по ошибкам регистра (T), при этом чаще применяет уместные трансформации, например, модуляцию: *“September was not prone to stealing, but neither was she entirely stupid”* – «Сентябрь не была склонна к воровству, но и совсем глупой ее назвать было нельзя».

- «Яндекс Переводчик» продемонстрировал наибольшую стабильность в переводе двух разных произведений. Это позволяет предположить, что глубина обученности переводчика на переводящем языке обеспечивает предсказуемость качества перевода, сокращая грубые ошибки эквивалентности, связанные с уровнем владения ПЯ.

Исходя из результатов проведенного исследования, мы можем дать несколько рекомендаций по улучшению качества работы нейронных машинных переводчиков при переводе художественных текстов. Так, на наш взгляд, необходимо:

- создавать специализированное обучение моделей на корпусах художественной литературы;
- внедрять динамические глоссарии для авторских терминов;
- совершенствовать механизмы сбора экспертных правок для дообучения систем.

Таким образом, развитие искусственного интеллекта в переводческой деятельности следует рассматривать не через призму конкуренции, а как возможность переосмысления профессии. Современные технологии НМП, несмотря на их когнитивные возможности, не способны автономно решать задачи, требующие креативной интерпретации и анализа. В этой связи, перспективной моделью представляется синергетическое взаимодействие в цепочке «НМП – переводчик – редактор», где нейросети выполняют функцию первичной обработки текста, а человек фокусируется на смысловой и стилистической коррекции.

ЛИТЕРАТУРА

1. Казакова Т. А. Художественный перевод: учебное пособие / Т. А. Казакова. – Санкт-Петербург : ИВЭСЭП, 2002. – 112 с. – ISBN 5-7320-0625-7.

2. Коваль О. И. Применение нейронных сетей для машинного перевода / О. И. Коваль, А. А. Петров // Новые импульсы развития: вопросы научных исследований. – 2020. – № 2. – С. 12–17. – URL: <https://cyberleninka.ru/article/n/primenenie-neuronnyh-setey-dlya-mashinnogo-perevoda> (дата обращения: 19.05.2025).
3. Лосева С. А. Перевод и искусственный интеллект: проблемы и пути развития / С. А. Лосева // Межкультурная коммуникация и профессионально ориентированное обучение иностранным языкам. – 2022. – № 12. – С. 162–168.
4. Морозкина Е. А. Использование информационных технологий для оптимизации процесса перевода / Е. А. Морозкина, Н. Р. Шакирова // Вестник Башкирского университета. – 2012. – № 1. – С. 544–546.
5. Мусоева Ш. Ю. Особенности художественного текста: специфика категории и функции / Ш. Ю. Мусоева. // Вестник Педагогического университета. – 2022. – № 3 (98). – С. 252–257.
6. Hinkel L. Hallucinating to better text translation. [Электронный ресурс] URL: https://news.mit.edu/2022/hallucinating-better-text-translation-0606?hss_channel=isoCknw3BEJc (дата обращения: 01.06.2025)
7. Adams, D. The Hitchhiker's Guide to the Galaxy / D. Adams. – New York : Random House, 2012. – 216 p. – ISBN 978-0-345-39180-3.
8. Valente C. M. The Girl Who Circumnavigated Fairyland in a Ship of Her Own Making / C. M. Valente. – London : Fairyland, 2012. – 288 p. – ISBN 978-1250010193.
9. Адамс Д. Автостопом по Галактике / перевод с английского В. И. Баканов // Д. Адамс. – Москва : АСТ, 2024. – 638 с. – ISBN 978-5-17-098748-1.
10. Валенте, К. М. Девочка, которая объехала Волшебную страну на самодельном корабле / перевод с английского В. Беленковича // К. М. Валенте. – Москва : АСТ, 2014. – 318 с. – ISBN 978-5-17-086775-2.
11. Matusov E. The Challenges of Using Neural Machine Translation for Literature. In Proceedings of the Qualities of Literary Machine Translation, pages 10–19, Dublin, Ireland. European Association for Machine Translation. 2019.
12. Смирнова А. И. Предвзятость как проблема алгоритмов ИИ: этические аспекты // Философия и общество. 2023 № 3 С. 118–126. DOI:10.30884/jfio/2023.03.07.
13. Харитонова Ю. С., Савина В. С., Паньини Ф. Предвзятость алгоритмов искусственного интеллекта: вопросы этики и права // Вестник Пермского университета. Юридические науки. 2021. Вып. 53. С. 488–515. DOI: 10.17072/1995-4190-2021-53-488-515.

УДК 81

Чуешкова Эльвира Ивановна

лаборант Высшей школы лингвистики
и педагогики

Санкт-Петербургский политехничес-
кий университет Петра Великого
г. Санкт-Петербург, Россия

Elvira Chueshkova

Laboratory Assistant of the Institute
of Humanites

Peter the Great St. Petersburg
Polytechnic University
St. Petersburg, Russia
elyachueshkova@gmail.com

ПРОБЛЕМЫ ЭКВИВАЛЕНТНОСТИ МЕДИЦИНСКИХ ТЕРМИНОВ В НЕЙРОННОМ МАШИННОМ ПЕРЕВОДЕ (на примере Google Translate и Yandex Translate)

Статья посвящена анализу эквивалентности англо-русского перевода медицинских терминов в нейронных машинных системах Google Translate и Yandex Translate. Материалом исследования послужили статьи журнала The Lancet (2025), из которых методом сплошной выборки отобраны 20 терминов различной структурной сложности. Переводы сопоставлялись по терминологической категории метрики MQM с учетом степени серьезности отклонений. Установлено, что обе модели обеспечивают в основном адекватную передачу понятий, однако допускают частично эквивалентные варианты, нарушают стилистическую сочетаемость и в отдельных случаях искажают значение. Google Translate чаще прибегает к калькированию и транскрибированию, приближая текст к международному стандарту, но увеличивая риск терминологических ошибок; Yandex Translate склонен к прагматической адаптации и использованию традиционных российских номинаций, снижая формальность, но повышая читабельность. Автор подчеркивает необходимость обязательного постредактирования специализированных текстов, совершенствования автоматических словарей и комплексных оценочных метрик. Результаты могут быть применены при доработке НМП-сервисов, а также в обучении медицинскому переводу и постредактированию.

Ключевые слова: нейронный машинный перевод; медицинская терминология; терминологическая эквивалентность; Google Translate; Yandex Translate.

CHALLENGES OF MEDICAL TERM EQUIVALENCE IN NEURAL MACHINE TRANSLATION (Google Translate and Yandex Translate)

The article focuses on analyzing the equivalence of English–Russian translations of medical terms produced by neural machine translation systems Google Translate and Yandex Translate. The research material includes articles from The Lancet (2025), from which 20 terms of varying structural complexity were selected using a continuous sampling method. Translations were compared based on the terminology category of the MQM metric, considering the severity of deviations. It was found that both models generally provide adequate conceptual rendering, yet produce partially equivalent variants, disrupt stylistic compatibility, and occasionally distort meaning. Google Translate tends to rely on calques and transliteration, aligning the text with international standards but increasing the risk of terminological errors. Yandex Translate, on the

other hand, leans toward pragmatic adaptation and the use of traditional Russian terms, reducing formality while improving readability. The author emphasizes the necessity of mandatory post-editing of specialized texts, the improvement of automatic dictionaries, and the development of comprehensive evaluation metrics. The results may be useful for improving NMT services as well as in training for medical translation and post-editing.

Key words: neural machine translation; medical terminology; terminological equivalence; Google Translate; Yandex Translate.

В условиях постоянного расширения и совершенствования человеческого знания наблюдается неуклонный рост числа новых терминов и языковых единиц, отражающих как переосмысление уже известных понятий, так и появление принципиально новых. Под влиянием глобализации происходит увеличение объемов международного сотрудничества в различных сферах деятельности.

Как известно, в последние годы широкое распространение получили цифровые инструменты, позволяющие автоматизировать и ускорить процесс перевода, редактирования и обработки текста. В числе таких инструментов сегодня доминирует нейронный машинный перевод (далее – НМП), пришедший на смену статистическим моделям машинного перевода, требовавшим значительных вычислительных ресурсов. Отличительной особенностью нейронных сетей выступает использование глубинного обучения (deep learning), направленного на достижение максимальной эффективности системы. Крупнейшие технологические компании, такие как Google, Яндекс, Microsoft и PROMT, уже интегрировали нейросети в свои сервисы перевода. Подобные технологии широко применяются в профессиональной переводческой практике, в частности, при работе с научными и техническими текстами. Тем не менее, даже самые передовые системы НМП продолжают допускать ошибки и неточности при передаче узкоспециализированной терминологии. Это создает необходимость тщательной проверки результатов работы нейронных переводчиков. Кроме того, существенно важной задачей представляется разработка систем оценки качества машинного перевода, а также совершенствование существующих метрик, поскольку данные технологии постоянно эволюционируют за счет продвинутых алгоритмов глубинного обучения и активного пополнения автоматических словарей. В рамках настоящего исследования рассматриваются модели двух популярных онлайн-сервисов перевода – Google Translate и Yandex Translate.

Материалом исследования послужили научные статьи в области медицины, опубликованные в журнале The Lancet за 2025 год. Этот ресурс представляет собой одно из самых авторитетных медицинских изданий в мире, которое освещает актуальные тенденции и открытия в сфере здравоохранения. Это делает данный корпус статей репрезентативным для анализа актуальной англоязычной медицинской терминологии. Современная медицинская литература отличается высокой плотностью и динамикой развития специализированной лексики: одни термины устаревают, другие – обретают

новые значения, третьи – только входят в обиход. Это связано с внедрением инновационных методов диагностики, лечения и новых медицинских технологий. Медицинская терминология представляет собой особый пласт языка, в котором точность формулировок и однозначность интерпретации имеют первостепенное значение. Малейшая неточность в переводе может повлечь за собой серьезные последствия – от искажения теоретической информации до ошибок в медицинской практике. На сегодняшний день в международном медицинском сообществе, как и в других отраслях, происходит постоянный обмен информацией и научным опытом, что требует качественного перевода и адаптации исследований на другие языки.

Таким образом, **цель** настоящего исследования заключается в анализе эквивалентности медицинских терминов при англо-русском нейронном машинном переводе с использованием онлайн-сервисов Google Translate и Yandex Translate.

Методология исследования включает в себя анализ научной литературы, касающейся особенностей функционирования медицинской терминологии, принципов нейронного машинного перевода и существующих подходов к оценке качества перевода. Также в исследовании используются методы лингвистического анализа, включая сопоставительный анализ переводов, а также качественный анализ оценки переводческих ошибок и несоответствий по системе MQM (Multidimensional Quality Metrics), с определением степени их серьезности.

Результаты проведенного исследования показали, что, несмотря на общую адекватность перевода терминов, модели Google Translate и Yandex Translate могут предоставлять низкоэквивалентные терминологические соответствия, а также нарушать стилистическую сочетаемость исходных и переведенных единиц. В связи с этим рекомендуется дополнительная проверка и коррекция переводов специализированных текстов, полученных при работе с системами нейронного машинного перевода. Описанные результаты могут быть использованы для оптимизации систем НМП, а также для обучения переводу и постредактированию текстов профессионального дискурса.

Для более точного понимания трудностей, которые представляет специализированный перевод, необходимо обратить внимание на определение термина как языковой единицы. В лингвистике существуют различные точки зрения относительно природы термина, его функций и места в лексической системе. Обобщая все аспекты этого явления, можно сказать, что термин – это специализированная языковая единица, предназначенная для точной номинации понятий профессиональной сферы.

Так, создатель российской терминологической школы Д. С. Лотте утверждает, что термином слово или словосочетание, обозначающее строго определенное понятие науки, техники или искусства [1, с. 8]. Также Лотте выделил у термина такие свойства, как отсутствие синонимов и омонимов в отдельной области, достаточная ясность, соответствие терминов родственных специальностей и способность термина вызывать определенную ассо-

цию. Согласно А. В. Суперанской, термины представляют собой «словесное обозначение понятия, входящего в систему понятий определенной области профессиональных знаний» [2, с. 14]. Термин понимается как вершина «понятийного дерева»; в основе этой идеи – строгое соотношение между понятием и знаком, который обозначает это понятие. Терминологическая система должна быть целостной и точно воспроизводить иерархию научных понятий. Иными словами, термины могут существовать только как компоненты единой лексической системы, которая, в свою очередь, формируется параллельно с расширением человеческих знаний в той или иной области.

В связи с высокими требованиями к организации и классификации терминов возросла необходимость унификации: «один термин – одно понятие». По мнению исследователей, установление нормы предотвращает многозначность и способствует преемственности научного знания. Поэтому для каждой науки и профессиональной области постоянно разрабатываются ГОСТы, ведомственные стандарты и терминологические словари. Благодаря работе российской терминоведческой школы были созданы первые отраслевые стандарты (ГОСТ 1.5-68, ГОСТ 7.0-84), методики терминологической работы при переводе, а также концепция «терминоэлементов» (*-cardio-, -itis, -scope*). На сегодняшний день подобная стандартизация значительно облегчает обмен опытом и информацией между сферами деятельности человека, а также способствует научному и техническому прогрессу.

Помимо облегчения классификации отраслевых терминологий, исследования в этой области также делают акцент на лингвокультурном аспекте термина. Международное научное и деловое сотрудничество предполагает, что все участники коммуникации должны пользоваться единой системой понятий для избежания неточностей и ошибок, особенно в высокопрофессиональной деятельности. Тем не менее, термин – это не просто носитель научного знания, но и отражение культурных, исторических и когнитивных особенностей сообщества, в котором он функционирует. Термины часто несут на себе отпечаток национальной картины мира, традиций научной школы, уровня развития отрасли и языка.

Поскольку термины рассматриваются как часть национального языка науки, при переводе крайне важна их адаптация к нормам словообразования и орфографии языка перевода. Помимо этого, одинаковые термины в разных языках могут иметь различные объемы значений или разную степень нормативности. Таким образом, при заимствовании терминологической единицы из одной системы знаний в другую невозможно не учитывать лингвокультурный фактор, поэтому интеграция международной лексики возможна лишь при учете системности и традиций языка реципиента [3].

В случае с медицинскими терминами, наиболее распространенным способом перевода является подбор словарных соответствий. Так, во многих случаях используется поиск единиц, эквивалентных по значению оригинальным. Чаще всего под этим подразумеваются названия, имена собственные,

а также лексические единицы с устойчивым значением, например, *neuropsychiatrist* ‘невропсихиатр’; *medical history* ‘история болезни’; *coronary arteries* ‘коронарные артерии’ [4, с. 124].

Принимая во внимание все особенности и функции терминов в научных текстах, необходимо рассмотреть принципы работы систем нейронного машинного перевода, в частности моделей Google и Yandex, при переводе узкоспециализированных текстов. Как известно, данные технологии используют обширные текстовые массивы для обучения, а внедрение нейросетей создает возможность обработки текста целиком, а не фрагментарно. Принцип подбора соответствий в каждой языковой паре функционирует благодаря созданию автоматических словарей, которые обеспечивают лексико-морфологический анализ обрабатываемого текста. Разработчики систем НМП используют цифровые таблицы для хранения и структуризации языковых единиц, что позволяет таким словарям расширяться и адаптироваться под новые технические условия. Однако стоит отметить, что автоматические словари не являются нормативными, поскольку в качестве заглавия словарной статьи используются все возможные варианты номинации объектов, а перевод соответствует рекомендуемому для конкретного языка перевода и предметной области [5, с. 11]. Следовательно, варианты перевода специальных терминов исходного текста, предлагаемые системами НМП, не всегда являются полностью эквивалентными и могут исказить значение результирующего текста перевода. На это необходимо обращать тщательное внимание при постредактировании, поскольку неточности в переводе медицинских исследований могут повлечь за собой серьезные последствия и отразиться на здоровье пациентов.

Сервисы Google и Yandex используют гибридные модели, совмещающие статистический машинный перевод на основе примеров и нейросетевые технологии, повышающие качество текста перевода. Модель Google Neural Machine Translation (GNMT) обучается на обширных параллельных корпусах, формируя словари на основе статистики употребления и контекста из этих корпусов. Yandex Translate основывается преимущественно на англо-русских и русско-английских массивах текстов. В этой системе используются как автоматические словари, так и частично ручная верификация лексем. Помимо этого, Yandex предоставляет возможность добавлять пользовательские варианты и быстрее реагирует на популярные запросы. Система может изменять приоритет словарных значений в зависимости от пользовательского выбора. Подобный принцип позволяет модели производить более естественные тексты, однако вовлечение пользователя может негативно сказаться на точности формулировок, в особенности, при переводе текстов научно-технической направленности.

На сегодняшний день существует множество систем оценки качества перевода, их появление обусловлено большими объемами информации, которыми необходимо обмениваться при осуществлении той или иной деятельности. Примером подобной системы служит метрика MQM (Multidimensional

Quality Metrics). Это унифицированная система оценки качества перевода, разработанная в рамках проекта QTLaunchPad при поддержке Европейской комиссии. Ее цель – предоставить гибкую, стандартизированную модель, позволяющую описывать, аннотировать и сравнивать качество перевода. Отличительной особенностью данной метрики выступает сравнение на основе экспертной оценки, а не только при помощи автоматических вычислений, на которых построены оценки BLEU, TER или BERTScore.

MQM представляет собой иерархическую систему категорий ошибок, которые могут встречаться в переводе, и предлагает подробную классификацию типов ошибок, их уровней серьезности и места возникновения. Она предназначена в первую очередь для ручной (экспертной) оценки качества перевода. MQM классифицирует ошибки по нескольким измерениям. Наиболее базовые категории включают **Meaning** (пропуски, добавление, искажение смысла), **Form** (орфография, пунктуация, грамматика), **Fluency** (стилистические ошибки), **Terminology** (неправильное использование терминов) и т. д. Также при работе с данной классификацией используется градация серьезности ошибок: **Minor** (незначительная) – не влияет на общий смысл, но нарушает стиль или правила; **Major** (значительная) – искажает часть смысла, может вызвать непонимание; **Critical** (критическая) – полностью искажает смысл, может повлечь серьезные последствия [6].

Метрика MQM может обеспечивать наиболее комплексную и объективную оценку качества перевода за счет привлечения экспертов, которые анализируют исходный и переведенный текст, после чего вручную аннотируют ошибки, руководствуясь детальной схемой. Это позволяет подсчитать общее число ошибок по категориям, определить их частотность и серьезность. Такой метод подходит для анализа человеческого, машинного и постредактированного перевода.

В настоящем исследовании были проанализированы и переведены фрагменты научных статей по медицине посредством Google Translate и Yandex Translate. Методом сплошной выборки были выделены специализированные термины, результат перевода которых отличался у двух используемых моделей НМП. Всего было отобрано 20 терминологических единиц, характеризующихся различной степенью сложности: от однокомпонентных терминов до сложных синтаксических конструкций и номинативных сочетаний.

Поскольку в рамках данного исследования рассматривается использование нейронного машинного перевода при работе с медицинской терминологией, расширенная оценка качества на основе MQM не представляется актуальной. Результаты перевода были сопоставлены по критерию терминологической эквивалентности (**Terminology**), то есть соответствия перевода понятийному содержанию оригинального термина с учетом его употребления в профессиональной среде. Стоит отметить, что при переводе и постредактировании узкоспециализированных текстов рекомендуется обращение к профессиональной литературе и словарям для выбора наиболее точного соответствия.

Анализ показал, что в большинстве случаев обе модели предлагают корректные терминологические эквиваленты:

- *analgesia* [7] ‘анальгезия’ (Google); ‘обезболивание’ (Yandex)
- *anticancer agents* [10] ‘противораковые средства’ (Google); ‘противоопухолевые средства’ (Yandex)
- *ileus* [9] ‘непроходимость кишечника’ (Google); ‘кишечная непроходимость’ (Yandex)

Эти случаи можно отнести к категории корректных специализированных соответствий, принятых в медицинской практике. Обе модели НМП предлагают синонимичные русскоязычные единицы, однако наблюдаются и случаи частичной эквивалентности:

- *carpal tunnel syndrome* [8] ‘синдром запястного канала’ (Google); ‘кистевой туннельный синдром’ (Yandex)

В данном примере переводчику или постредактору необходима тщательная проверка терминов, поскольку лексемы, связанные с диагностикой или назначением лечения, требуют максимально точной передачи значения.

Следующие примеры демонстрируют стилистические различия между вариантами перевода, предлагаемыми двумя моделями:

- *revision surgery* [7] ‘ревизионная операция’ (Google); ‘повторная операция’ (Yandex)
- *sedation* [9] ‘седация’ (Google); ‘седативный эффект’ (Yandex)
- *hypersalivation* [9] ‘гиперсаливация’ (Google); ‘повышенное слюноотделение’ (Yandex)

Как мы можем видеть, система Google предоставляет более специализированные, научно принятые номинации, в то время как Yandex предлагает более обиходные, но естественно звучащие конструкции. Это означает, что в первом случае модель НМП часто переводит термины через транскрибирование, что делает переведенный текст более подходящим под международные научные стандарты, тогда как вторая система более дескриптивна и ориентирована на среднестатистического читателя. Данное различие между моделями отражает прагматическую направленность результирующего текста, что можно принять во внимание, исходя из требований к переводу.

Также следует отметить, что автоматические словари Yandex Translate более адаптированы к российской медицинской традиции. Это демонстрируется в следующих примерах:

- *ankylosing disease* [7] ‘анкилозирующая болезнь’ (Google); ‘болезнь Бехтерева’ (Yandex)
- *primary care physicians* [8] ‘врачи общей практики’ (Google); ‘врачи первичной медицинской помощи’ (Yandex)
- *occupational therapists* [8] ‘эрготерапевты’ (Google); ‘трудотерапевты’ (Yandex)

- *antipsychotics* [9] ‘антипсихотические препараты’ (Google); ‘нейролептики’ (Yandex)

Несмотря на корректность обоих вариантов перевода, модель Yandex чаще предлагает термины, закреплённые в российском медицинском сообществе, что делает переведённый текст более доступным для русскоязычных специалистов.

Как известно, модели НМП не всегда способны учитывать исходный контекст и ориентироваться на него при подборе словарных соответствий:

- *titration* [9] ‘титрование’ (Google); ‘повышение дозы’ (Yandex)

В этом примере система Google предложила вариант перевода с низкой эквивалентностью, поскольку понятие «титрование» является многозначным и может означать «метод химического анализа» в других контекстах. Таким образом, некорректный выбор соответствия может отразиться на интерпретации исходного текста:

...weight gain and sedation often lead to discontinuation; these can be mitigated by slower titration or augmentation with a less-sedating antipsychotic [9].

Таким образом, обе модели НМП демонстрируют высокую вариативность при подборе терминологических эквивалентов, однако и Google, и Yandex обладают как преимуществами, так и недостатками. Анализ ошибок и неточностей проводился на основе категории Terminology по метрике MQM, с фокусом на искажение терминов, недостаточную специфичность или стилистическую некорректность. Результаты проведенного анализа позволяют сделать несколько выводов.

- Модель Google Neural Machine Translation (GNMT) в большинстве случаев демонстрирует адекватную передачу медицинских терминов, близких к международным стандартам. Несмотря на это, Google бывает склонен к калькированию, а также к буквальному переводу сложных конструкций (*long-segment posterior pedicle-screw surgical fixation* [7] ‘фиксация длинным сегментом заднего транспедикулярного винта’). В связи с этим система часто предоставляет некорректные соответствия, которые могут привести к критическим ошибкам в переведённом тексте.

- Yandex Translate предлагает пользователю более общеупотребимые варианты перевода и более лаконичные формулировки для сложных терминов. Это означает, что результат перевода может оказаться менее формальным и нормированным, чем исходный текст, что является стилистической ошибкой. Тем не менее, Yandex может демонстрировать более эффективную прагматическую адаптацию оригинала.

- Оба сервиса нуждаются в тщательном постредактировании при работе с узкоспециализированными текстами.

Особую важность терминологическая точность приобретает в контексте перевода медицинских текстов, где термин не просто обозначает объект или процесс, но и зачастую несет в себе информацию, критически значимую

для диагностической, терапевтической или исследовательской деятельности. Медицинские термины выполняют не только номинативную, но и когнитивную функцию: они являются носителями строго определенного понятийного содержания, закреплённого в профессиональном дискурсе.

В условиях активного внедрения технологий нейронного машинного перевода возрастают риски автоматической подстановки терминологических соответствий без учета контекста, жанровой принадлежности текста и лингвокультурных особенностей языка перевода. Таким образом, на сегодняшний день проверка и коррекция терминов, предложенных системами НМП, представляют собой неотъемлемую часть профессиональной переводческой деятельности.

ЛИТЕРАТУРА

1. Лотте Д. С. Основы построения научно-технической терминологии. Вопросы теории и методики. М. : Изд-во АН СССР, 1961. 160 с.
2. Суперанская А. В., Подольская Н. В., Васильева Н. В. Общая терминология: вопросы теории. М. : ЛИБРОКОМ, 2012. 246 с.
3. Лейчик В. М. Терминоведение. Предмет, методы, структура. 4-е изд. М. : Книжный дом «ЛИБРОКОМ», 2009. 256 с.
4. Комиссаров В. Н. Лингвистика перевода. М. : Международные отношения, 2013. 353 с.
5. Беляева Л. Н., Камшилова О. Н. Лексикографические проблемы систем машинного перевода: на пути от буквального до нейронного // Вестник ВолГУ. Серия 2: Языкознание. 2024. № 5.
6. TheMQM.org [Электронный ресурс]. URL: <https://themqm.org/> (дата обращения: 05.07.2025).
7. Beucle N. The burden of unstable fractures in axial spondyloarthritis // The Lancet, 2025.
8. Zimmerman M., Dahlin L. B. New evidence for common practices in carpal tunnel syndrome // The Lancet, 2025.
9. Fanshawe J. B., Kabir T., Lennox B. R. Clozapine: old drugs require efforts to improve patient experience and access // The Lancet, 2025.
10. Heinrich M. Medicines from the garden // UCL School of Pharmacy (Pharmacognosy & Phytotherapy Research Group), University College London // The Lancet, 2025.

Секция 6.
МЕЖДИСЦИПЛИНАРНЫЕ ИССЛЕДОВАНИЯ
НА ОСНОВЕ АНАЛИЗА ТЕКСТОВЫХ ДАННЫХ

УДК 81'322.2:81'36:811.161.3

Варановіч Валерый Віктаравіч
старшы выкладчык
Беларускі дзяржаўны ўніверсітэт
г. Мінск, Беларусь

Valery Varanovich
Senior Lecturer
Belarusian State University
Minsk, Belarus
gamrat.vvv@gmail.com

Гецэвіч Юрась Станіслававіч
кандыдат тэхнічных навук,
AI распрацоўшчык
ТАА “СОФТКЛУБ – Цэнтар
распрацоўкі”
г. Мінск, Беларусь

Yuras Hetsevich
PhD in Science,
AI research & systems integrator
JSC SoftClub
Minsk, Belarus
yuras.hetsevich@gmail.com

Супрунчук Мікіта Віктаравіч
кандыдат філалагічных навук,
дацэнт
Беларускі дзяржаўны ўніверсітэт
замежных моў
г. Мінск, Беларусь

Mikita Suprunchuk
PhD in Philology,
Associate Professor
Belarusian State University
of Foreign Languages
Minsk, Belarus
suprunchuk@mail.ru

АЎТАМАТЫЗАВАНАЯ АПРАЦОЎКА БЕЛАРУСКАЙ МОВЫ
ПРАЗ КАМП'ЮТАРНА-ЛІНГВІСТЫЧНЫ МОДУЛЬ NOOJ

Каб атрымаць некаторыя звесткі лінгвістычнага (пра мову) ці экстралінгвістычнага (пра свет) характару, трэба вывучаць вялікія тэксты, бо апора на дастатковую колькасць фактаў звычайна забяспечвае пэўную і дакладную інфармацыю. У гэтым дапамагаюць метады корпуснай лінгвістыкі і аўтаматычнай апрацоўкі мовы (NLP). Для аблягчэння стварэння карпусоў і далейшай падрыхтоўкі запытаў да іх выкарыстоўваюцца анлайнавыя і афлайнавыя корпусныя менеджары. Прыкладам апошняга з іх з'яўляецца камп'ютарны лінгвістычны працэсар NooJ, які на пачатку 2000-х гг. распрацаваў М. Зільберштайн (Францыя). Беларускі модуль працэсара быў падрыхтаваны ў 2012 г. У дакладзе апісаны даследаванні з яго выкарыстаннем і асобныя дапаўненні да яго. Апісана падрыхтоўка новага рэпрэзентатыўнага міні-корпуса ў 1,5 мільёна словаўжыванняў. У беларускім базавым слоўніку выпраўлена тэгіраванне дзеепрыслоўяў, прапанаваны сінтаксічныя правілы для больш дакладнага аўтаматычнага тэгіравання амонімаў, для вылучэння і аднастайнай апрацоўкі фразеалагізмаў; дададзена каля 300 новых слоў і дапоўнена граматычная інфармацыя пра асобныя словы.

Ключавыя словы: аманімія; аўтаматычная апрацоўка мовы; корпус тэкстаў; корпусны менеджар; сінтаксічная граматыка; NLP; NooJ.

AUTOMATED PROCESSING OF THE BELARUSIAN LANGUAGE WITH NOOJ COMPUTER-LINGUISTIC MODULE

To obtain some information of a linguistic (about language) or extralinguistic (about the world) nature, it is necessary to study large texts, since reliance on a sufficient number of facts usually provides certain and accurate information. Methods of corpus linguistics and automatic language processing (NLP) help in this. Online and offline corpus managers are used to facilitate the creation of corpora and the further preparation of queries for them. The computer linguistic processor NooJ is an example of the latter. It was developed by M. Silberstein (France) in the early 2000s. The Belarusian module of the processor was prepared in 2012. The article describes research done with it and some recent additions to it. The preparation of a new representative mini-corpus of 1.5 million word usages is described. The Belarusian basic dictionary has been corrected (the tagging of gerunds), syntactic rules for more accurate automatic tagging of homonyms and for uniform picking out and processing of phraseological units were proposed; and 300 new words and supplemented grammatical information about individual words were added.

Key words: corpus manager; homonymy; NLP; NooJ; text corpus; syntactic grammar.

1. Некаторыя сродкі аўтаматызаваанай апрацоўкі натуральнай мовы.

Даследаванні мовы, структуры і арганізацыі ведаў зараз актыўна вядуцца з прымяненнем камп'ютарных сродкаў. Вялікую дапамогу пры аналізе і апісанні моўных з'яў аказваюць электронныя карпусы і праграмы аўтаматычнай апрацоўкі тэксту. У інтэрнэце выстаўлены вялікія карпусы тэкстаў на розных мовах, якія адкрываюць шырокія магчымасці вывучэння семантыкі і спалучальнасці адзінак розных моўных узроўняў. Пры гэтым трэба ўлічваць, што ў большасці выпадкаў для пошуку даступны загаловак загрузаныя і падрыхтаваныя творы, а новыя (свае) матэрыялы нельга дадаць у агульнадаступны корпус, напрыклад у нацыянальны. Таму пры неабходнасці аналізу дадатковых, не ўключаных у корпус тэкстаў выкарыстоўваюцца спецыяльныя праграмы – корпусныя менеджары. Адным з удалых прыкладаў такіх праграм з'яўляецца моўны працэсар NooJ [1].

NooJ быў распрацаваны, абапіраючыся на вынікі даследаванняў прафесара Макса Зільберштайна і калектыву INTEx – аб'яднання лінгвістаў, якія прытрымліваюцца лексіка-граматычнай метадалогіі, уведзенай Марысам Гросам ў Лабараторыі аўтаматызаваанай дакументацыі і лінгвістыкі. Гэта метадалогія зыходзіць з таго, што любое граматычнае правіла патрабуе выразнага вызначэння сферы яго прымянення, акцэнтуючы значнасць кантэксту ў працэсе лінгвістычнага разбору. Гэта дазваляе лінгвістам ствараць і аналізаваць усе чатыры класы граматык (паводле Н. Хомскага): рэгулярныя, кантэкстна-свабодныя, кантэкстна-залежныя і неабмежаваныя граматыкі. Карыстальнікі могуць ствараць іх, выкарыстоўваючы альбо тэкставы рэдактар для напісання рэгулярных выразаў, альбо графічны рэдактар для візуальнага прадстаўлення.

Праблема ўдасканалення электронных тэкстаў, напісаных на натуральнай мове, таксама вырашаецца з дапамогай праграмнага комплексу NooJ. Ён

дае магчымасць рыхтаваць сінтаксічныя і марфалагічныя правілы (алгарытмы, граматыкі) у выглядзе канчатковых аўтаматаў. Акрамя таго, NooJ забяспечвае зручнае тэсціраванне эфектыўнасці такіх граматык на тэкстах.

2. Беларускі модуль Нудж: аснова і кірункі абнаўлення.

Працэсар NooJ уяўляе сабой праграмную абалонку, якая дазваляе ствараць, рэдагаваць і захоўваць тэкставыя файлы, рабіць некаторыя статыстычныя падлікі. Для карэктнай работы да яе неабходна падключаць адпаведны моўны модуль. Модуль звычайна складаецца з граматычнага слоўніка (набор слоў з іх парадыгмамі), спіса граматычных катэгорый і значэнняў, аднаго ці некалькіх тэкстаў-узораў, алгарытмаў-правіл для паляпшэння лінгвістычнай разметкі. Модуль для беларускай мовы ў 2011–2012 гг. распрацавалі Юрась і Святлана Гецэвічы (Аб'яднаны інстытут праблем інфарматыкі НАН Беларусі) ды Яўгенія Якубовіч (Аўтаномны ўніверсітэт Барселоны, Іспанія). Гэты кампанент уключае ў сябе набор скампіляваных тэкстаў, некалькі дэманстрацыйных граматык, а таксама набор слоўнікаў агульным аб'ёмам каля 137 тыс. слоў. Своеасаблівым дадаткам гэтага модуля можна лічыць электронны корпус тэкстаў [2], падрыхтаваны ў 2015 г.

Названыя камп'ютарныя сродкі паспяхова выкарыстоўваліся для даследавання самых розных аспектаў беларускай, рускай і англійскай моў. Выявілася, што гэта вельмі зручны інструмент для стварэння карпусоў, паўаўтаматычнага пошуку слоў, выразаў, канструкцый, добрая аснова пры падрыхтоўцы тэкстаў да аўтаматычнага агучвання (сінтэзу маўлення). Разам з тым за мінулыя 10 гадоў былі знойдзены і пэўныя недахопы ці прапанаваны новыя, лепшыя рашэнні.

3. Удасканаленне лексічнай часткі беларускага модуля.

Пры стварэнні і апрацоўцы карпусоў юрыдычных, медыцынскіх і сацыяльных тэкстаў [3; 4] мы выявілі пэўную колькасць невядомых слоў. Да іх адносяцца:

1. Словы і словаформы, якія не ўключаны ў агульны слоўнік `dictionary_be.pod`: *апека, ільготны*.

2. Англійскія словы і сімвалы: *http, documents, e-mail*.

3. Рымскія лічбы: *V, X*.

4. Звычайныя і друкарскія памылкі: *стацілі, праналежнасць*.

5. Словы з апострафам: *з'езд, аб'яднаць*. Нязручная асаблівасць NooJ ў тым, што яго алгарытмы ўспрымаюць апостраф як мяжу паміж словамі. Каб гэтага пазбегнуць, трэба кожны выпадак маркіраваць у слоўніку спецыяльнай пазнакай (+UNAMB).

Даволі нечакана аказалася, што ў слоўніку не хапае пэўнай колькасці слоў, як агульнаўжывальных, так і тэрмінаў. Таму было вырашана дапоўніць слоўнік такімі агульнаўжывальнымі словамі, як *капітал, каштоўны, квартал, прастора* і інш., – усяго 213, з іх 106 назоўнікаў, 56 прыметнікаў, 16 дзеясловаў і інш. Дададзены былі і тэрміны: *апекавальнік, радыёчастата, чартар* і інш.

Зразумела, што неабходна было дадаваць не толькі пачатковую форму слова, але і іншыя. Для падбору і генерацыі парадыгм выкарыстоўваўся сэрвіс The Word Paradigm Generator [5], які прапаноўваў некалькі варыянтаў марфалагічнага ўзора (гл. малюнак 1).

Наступная задача, якая вырашалася, звязана з фразеалагізмамі. Было неабходна распрацаваць алгарытмы, якія адрозніваюць свабодныя словазлучэнні ад звязаных (фразеалагізмаў). Для гэтага былі падрыхтаваны графічныя правілы (візуальныя сінтаксічныя графы). Яны дазваляюць вылучаць фразеалагізмы адпаведна слоўніку І. Я. Лепешава [6] і забяспечваюць правільную граматычную разметку, такім чынам, з'яўляюцца добрай базай для аўтаматычнага сінтэзу маўлення і аўтаматычнага перакладу.

computational text & speech corpus by Word Paradigm Generator

Please, enter some wordforms from paradigm

апекавальнік,NOUN
апекавальніка,NOUN
апекавальніку,NOUN

☐ Processing according to wordforms dictionary
☒ Processing according to dictionary of inflections in NooJ format

Generate probable paradigms!

Result

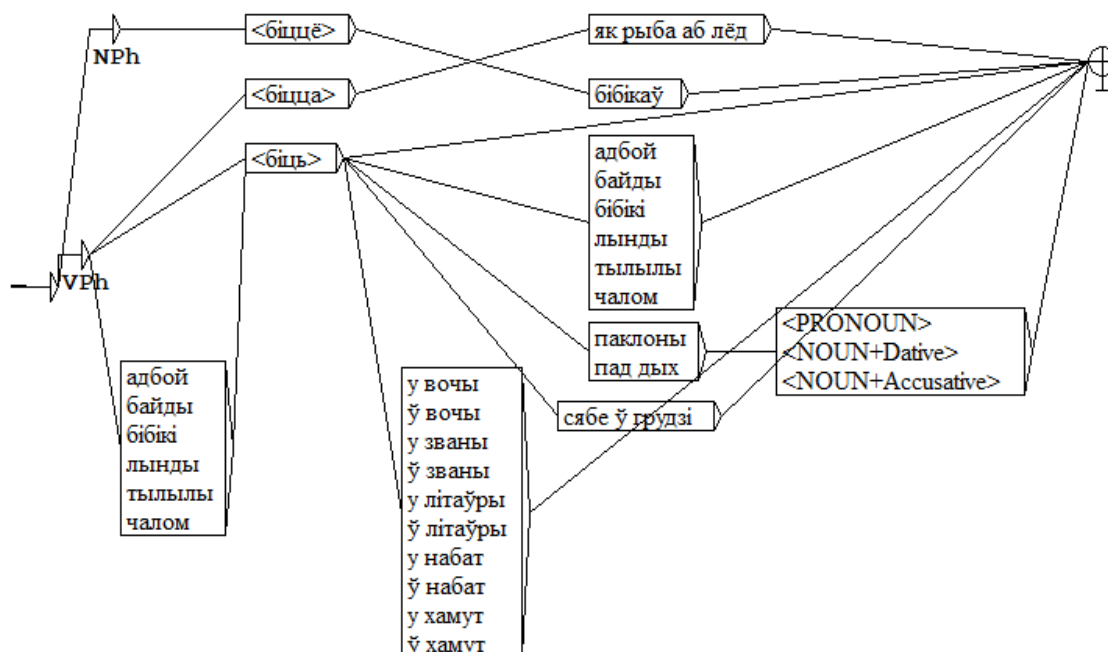
Paradigms are based on 3 forms (Total 152):

апекавальнік,NOUN+FLX=АБАГАЧАЛЬНІК
апекавальнік/Accusative+Common+Inanimate+Masculine
апекавальнік/Common+Inanimate+Masculine+Nominative
апекавальнікі/Accusative+Common+Inanimate+Masculine+Plural
апекавальнікі/Common+Inanimate+Masculine+Nominative+Plural
апекавальніка/Common+Genitive+Inanimate+Masculine
апекавальнікаў/Common+Genitive+Inanimate+Masculine+Plural
апекавальнікам/Common+Inanimate+Instrumental+Masculine
апекавальнікам/Common+Dative+Inanimate+Masculine+Plural
апекавальнікамі/Common+Inanimate+Instrumental+Masculine+Plural
апекавальніках/Common+Inanimate+Masculine+Plural+Prepositional
апекавальніку/Common+Dative+Inanimate+Masculine
апекавальніку/Common+Inanimate+Masculine+Prepositional;

Малюнак 1. Сэрвіс парадыгм The word paradigm generator

Напрыклад, дзякуючы граматыцы на малюнку 2 праграма будзе аўтаматычна аб'ядноўваць слова *біць* і наступныя ў фразеалагізм, лічыць іх адной адзінкай, што, натуральна, уплывае на вымаўленне і пераклад выказаў.

Belarusian/Belarusian syntactic grammar.



Малюнак 2. Алгарытм выдзяленне фразеалагізмаў са словам *біць*

4. Удасканаленне граматычнай часткі беларускага модуля.

Асноўныя складанасці ў граматычнай частцы беларускага модуля можна падзяліць на дзве групы: а) разнастайныя недакладнасці граматычнай разметкі; б) няправільная апрацоўка дзеепрыслоўяў. Да першай групы адносіцца, напрыклад, няправільная кваліфікацыя слова *павінен* (як дзеепрыметнік замест прыметніка) і некаторых іншых, а таксама недакладнае вызначэнне асобных формаў ступені параўнання прыметнікаў. Часам памылкова як два словы апрацоўваюцца вытворныя прыназоўнікі, напрыклад *на працягу*, *за выключэннем*.

Таксама выяўлена, што дзеепрыслоўе няправільна анатуецца ў беларускіх тэкстах. У славянскім мовазнаўстве гэтая форма звычайна адносіцца да дзеяслоўнай парадыгмы (нязменная форма дзеяслова). Аднак пры стварэнні флексійнага слоўніка NooJ дзеепрыслоўе было вылучана як асобны клас (часціна мовы), што прывяло да праблем пры яго апрацоўцы.

На самай справе праблем было нават дзве. Па-першае, некаторыя формы дзеясловаў пры анатаванні тэкстаў таксама вызначаліся як дзеепрыслоўі, што з’яўляецца памылкай. Па-другое, некаторыя дзеепрыслоўі не анатаваліся належным чынам.

У выніку, паўстала неабходнасць выправіць набор тэгаў “Verb” і перакампіляваць слоўнік *general_be.dic*. У галоўным спісе тэгаў былі зробленыя наступныя змены: выдалены файл з тэгамі дзеепрыслоўяў, абноўлены файл тэгсэта дзеяслова. Пасля быў запушчаны алгарытм канвертацыі файлаў галоўнага слоўніка ў файлы ў фармаце NooJ. У выніку галоўныя файлы слоўніка NooJ былі атрыманыя ў выглядзе *_properties.def*, *general_be.dic*, *general_be.nof* (малюнак 3).

```

АБАЊІРАВАЦЬ =
<B5>аваў/Conjugated+Indicative+Masculine+Past+Perfective+Singular
+ <B5>аваўшы/Gerund+Perfective
+ <B5>авалі/Conjugated+Past+Perfective+Plural
+ <B5>авала/Conjugated+Feminine+Indicative+Past+Perfective+Singular
+ <B5>авала/Conjugated+Neuter+Past+Perfective+Singular
+ <B5>аваць/Perfective+infinitive
+ <B5>уе/3+Conjugated+Future+Indicative+Perfective+Singular
+ <B5>уем/1+Conjugated+Future+Indicative+Perfective+Plural
+ <B5>уеце/2+Conjugated+Future+Indicative+Perfective+Plural
+ <B5>уеш/2+Conjugated+Future+Indicative+Perfective+Singular
+ <B5>уй/Conjugated+Imperative+Perfective
+ <B5>уйце/Conjugated+Imperative+Perfective
+ <B5>ую/1+Conjugated+Future+Indicative+Perfective+Singular
+ <B5>уюць/3+Conjugated+Future+Indicative+Perfective+Plural;

```

Малюнак 3. Выпраўленне дзеяслоўнай парадыгмы (general_be.nof)

Таксама былі вывучаны заканамернасці ўжывання часавых форм дзеяслова ў залежнасці ад кантэксту, зафіксаваны амографы і амаформы, якія ўзнікаюць пры супадзенні форм цяперашняга часу, інфінітыва, імператыва (лячу, расці, закапаць і інш.), дзеяслова і іншых часцін мовы (ніла, мыла, вяла, складаў і інш.).

5. Лексіка-граматычныя праблемы.

Шэраг недакладнасцей аўтаматычнай разметкі звязаны з багатай аманічнасцю слоў і асобных форм у беларускай мове. Падчас практычнай дзейнасці былі выяўлены і поўныя (супадзенне ўсіх форм парадыгмы), і частковыя амонімы (амаформы ці амографы). Значыць, пільнай увагі патрабавала праблема аманіміі (неадназначнасці).

У сучаснай камп'ютарнай лінгвістыцы можна вылучыць чатыры ключавыя падыходы да вырашэння пытання няяснага значэння слова:

1) метады навучання “з настаўнікам” выкарыстоўваюць маркіраваныя тэкставыя карпусы для навучання класіфікатараў (supervised methods) [7].

2) метады частковага навучання “з настаўнікам” грунтуюцца на дадатковых ведах, такіх як азначэнні тэрмінаў у тлумачальных слоўніках або паралельныя двухмоўныя карпусы (minimally-supervised methods) [6];

3) метады навучання “без настаўніка” не патрабуюць выкарыстання знешніх крыніц інфармацыі і працуюць выключна з неанатаванымі карпусамі (сырымі немаркіраванымі карпусамі). Яны таксама вядомыя як кластарызацыя і размежаванне семантычных значэнняў слоў (unsupervised methods) [9].

4) метады, якія абапіраюцца на веды, у асноўным звяртаюцца да слоўнікаў, тэзаўрусаў і лексікаграфічных рэсурсаў (knowledge-based and dictionary-based methods) [10, с. 110–111]. Падрабязней гл. [11].

Як паказалі нашы папярэднія даследаванні, у беларускіх тэкстах неадназначнасць дасягае 50 % [12]. На працягу 10 гадоў супрацоўнікі Лабараторыі

сінтэзу і распазнавання маўлення Аб'яднанага інстытута праблем інфарматыкі НАН Беларусі вывучалі спосабы яе змяншэння. Для бягучага абнаўлення беларускага модуля NooJ было абрана 50 алгарытмаў (правіл), якія дадаюцца да папярэдняй версіі модуля і дапамагаюць праграме тэгіраваць тэкст.

Парадак пабудовы правіл (граматык) для вырашэння аманіміі на аснове эксперыментальнага корпуса быў наступны.

1. Была сабраная агульная статыстыка па розных відах аманіміі з дапамогай інструментаў NooJ.

2. Сярод усіх выпадкаў аманіміі былі вылучаны выпадкі граматычнай аманіміі.

3. Далей яны былі класіфікаваныя па часцінах мовы.

4. Потым быў створаны частотны слоўнік аманімічных словаформ.

5. Найбольш частотныя выпадкі граматычнай аманіміі аналізаваліся экспертамі-лінгвістамі і рабілася выснова, ці можна зняць гэту аманімію з дапамогай кантэкстуальнай граматыкі NooJ.

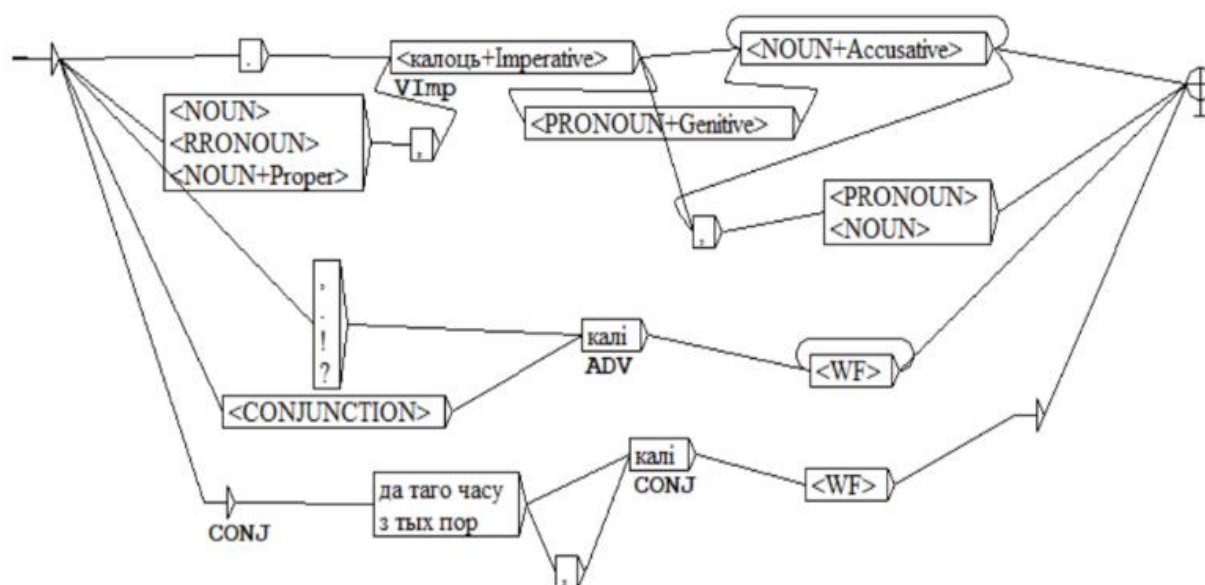
6. Наступны крок – складанне канкардансаў для выбраных выпадкаў з мэтай аналізу кантэкстаў, якія дапамогуць вырашыць аманімію.

7. На аснове канкардансаў ствараліся граматыкі з указаннем найбольш характэрных кантэкстаў для пэўных словаформ. На гэтым этапе граматыкі прымяняліся да названага вышэй эксперыментальнага корпуса і фіксавалася колькасць выпадкаў, якія граматыка пакрывае (здзімаецца аманімія).

8. Вылічаўся працэнт выпадкаў, пакрытых граматыкай, у адносінах да агульнай колькасці ўжывання ў корпусе пэўнай словаформы.

9. Граматыка лічылася эфектыўнай і ўвайшла ў канчатковы склад набору граматык, калі працэнт пакрывальнасці быў вышэйшы за 50 %.

Прыклад граматыкі пададзены на малюнку 4.



Малюнак 4. Сінтаксічная граматыка для вырашэння аманіміі словаформы *калі*

Граматыка дазваляе, арыентуючыся на кантэкст, вызначыць, з’яўляецца словаформа *калі* злучнікам, прыслоўем ці імператывам ад дзеяслова *калоць*. Звернем увагу, што вырашыць гэту аманімію дапамагаюць не толькі суседнія словы, але і знакі прыпынку.

6. Заключэнне.

У наш час перад даследчыкамі мовы ўзнікаюць новыя, усё больш складаныя задачы. Існуе патрэба аналізаваць вялікія аб’ёмы тэкстаў і пры гэтым вышукваць больш падрабязную інфармацыю пра лексічнае і граматычнае значэнне, спалучальнасць. Значнае месца ў працы лінгвіста цяпер займае праца з карпусамі, і ў гэтым дапамагаюць сродкі аўтаматызацыі, у тым ліку корпусныя менеджары. Гэты артыкул знаёміць з магчымасцямі праграмы NooJ, коратка прэзентуе гісторыю яго стварэння, развіццё беларускага дадатку да яго і нядаўнае абнаўленне. Удасканаленне беларускага модуля NooJ было накіравана на папаўненне слоўніка агульнай і тэрміналагічнай лексікай, удакладненне інфармацыі пра парадэгмы слоў, на змену падачы граматычных класаў слоў, на распрацоўку сінтаксічных правіл (граматык) для змяншэння недназначнасці (аманіміі). Гэта дазволіла палепшыць якасць анатавання тэкстаў, зрабіла слоўнік і граматыку NooJ больш адпаведнымі беларускаму маўленню. У планах распрацоўшчыкаў поўнае абнаўленне граматычнага набору (тэксэта), дадаванне фанетычнай інфармацыі ў слоўнік і інш. Гэта дапаможа палепшыць якасць сістэм аўтаматычнага перакладу і сінтэзу маўлення.

ЛІТАРАТУРА

1. Silberstein M. NooJ: a corpus processor [Electronic resource]. URL: <https://nooj.univ-fcomte.fr/> (date of access: 19.06.2025).
2. The first one-million corpus for the Belarusian NooJ module / I. Reentovich, Y. Hetsevich, V. Voronovich [et al.] // Automatic processing of natural-language electronic texts with NooJ: 9th International conference, Minsk, Belarus, June 11–13, 2015: revised selected papers / eds. T. Okrut, Y. Hetsevich, M. Silberstein, H. Stanislavenka. Springer International Publishing, 2016. P. 3–15.
3. Creation of a legal domain corpus for the Belarusian module in NooJ: texts, dictionaries, grammars / V. Varanovich, M. Suprunchuk, Y. Zianouka [et al.] // 16th International conference, NooJ 2022 / Universidad Nacional de Rosario ; ed. M. González [et al.]. Rosario : Springer Cham, 2023. P. 151–162.
4. Creation of parallel medical and social domains corpora for the machine translation and speech synthesis systems / M. Suprunchuk, N. Yarash, Y. Hetsevich [et al.] // 16th International conference, NooJ 2022 / Universidad Nacional de Rosario ; ed. M. González [et al.]. Rosario : Springer Cham, 2023. P. 139–150.
5. The word paradigm generator [Electronic resource]. URL: <https://corpus.by/wordparadigmgenerator> (date of access: 19.06.2025).
6. Лепешаў І. Я. Этымалагічны слоўнік фразеалагізмаў. Мінск : БелЭн, 2004. 448 с.

7. Word sense disambiguation: algorithms and applications / eds. E. Agirre, P. Edmonds. Springer, 2007. 377 p.
8. Van den Beukel S., Aroyo L. Homonym detection for humor recognition in short text // Proceedings of the 9th Workshop on computational approaches to subjectivity, sentiment and social media analysis. Brussels, Belgium, 31 October 2018. Brussels, Belgium, 2018. Pp. 286–291.
9. Зинькина Ю. В., Пяткин Н. В., Невзорова О. А. Разрешение функциональной омонимии в русском языке на основе контекстных правил // Труды междунар. конф. Диалог'2005. М. : Наука, 2005. С. 198–202.
10. Navigli R., Velardi P. Structural semantic interconnections: a knowledge-based approach to word sense disambiguation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005. Vol. 27, no. 7. Pp. 1075–1086.
11. Мадэль аўтаматызаваанай ідэнтыфікацыі амографаў для беларускай мовы / Ю. С. Гецэвіч, Я. С. Зяноўка, Д. І. Латышэвіч [і інш.] // Інформатика. 2023. Т. 20, № 4. С. 87–100.
12. Automatic grammatical disambiguation in Belarusian and Russian legal domain / M. Suprunchuk, V. Varanovich, Y. Zianouka, Y. Hetsevich // Formalizing natural languages: applications to natural language processing and digital humanities, NooJ 2024. Communications in computer and information science, vol. 2443 / ed. M. Silberztein. Springer, Cham, 2025. DOI: https://doi.org/10.1007/978-3-031-89810-5_7. Pp. 79–90.

УДК 811.11

Власова Диана Валентиновна

аспирант

Российский университет дружбы народов

г. Москва, Россия

Diana Vlasova

Postgraduate student

RUDN University

Moscow, Russia

vlasova_dv@pfur.ru

ПСИХОЛИНГВИСТИЧЕСКИЕ ОСНОВЫ ИЗУЧЕНИЯ АНГЛОЯЗЫЧНОГО РЕЧЕВОГО ЭТИКЕТА В СОВРЕМЕННЫХ КОРПУСНЫХ ИССЛЕДОВАНИЯХ

Данная статья посвящена возможности продуктивного изучения языковых формул англоязычного речевого этикета в контексте совместного функционирования психолингвистики и корпусных исследований. Особое внимание уделяется когнитивным и психолингвистическим механизмам воспроизведения и последующего восприятия этикетных формул, а также возможности их анализа на основе эмпирических данных крупных корпусов английского языка. Предлагается теоретическое обоснование направлений анализа функционирования средств речевого этикета в структуре естественного речевого акта с акцентом на их роли в реализации интенциональной направленности высказывания. Представляется, что интеграция психолингвистического и корпусного подходов позволяет глубже понять речевое поведение говорящего и способствует эффективному межкультурному общению, которое строится на соответствии нормативным моделям коммуникативного поведения в рамках повседневного дискурса. Объединение психолингвистического и корпусного подходов позволяет не только описывать речевой этикет как систему, но и анализировать его функционирование в динамике, на основании репрезентативных данных.

Ключевые слова: англоязычный речевой этикет; психолингвистика; корпусная лингвистика; когнитивные механизмы; механизмы речепорождения.

PSYCHOLINGUISTIC FOUNDATIONS OF STUDYING ENGLISH SPEECH ETIQUETTE IN MODERN CORPUS RESEARCH

The article discusses the potential for a productive study of linguistic formulas in English-language speech etiquette within the context of psycholinguistic and corpus-based research. It focuses on the cognitive and psycholinguistic processes involved in the production and perception of these etiquette formulas, as well as on the possibilities for their analysis using empirical data from large-scale English language corpora. The article proposes a theoretical framework for analyzing the use of speech etiquette in natural speech acts, highlighting their role in conveying the speaker's intentions. It seems that the integration of psycholinguistic and corpus methods allows for a better understanding of the speaker's speech patterns and promotes effective cross-cultural communication, which is based on following normative models of communication within the context of everyday discourse. By combining these two approaches, it is possible not only describe speech etiquette as a system, but also analyze its dynamics based on representative data.

Key words: English-language speech etiquette; psycholinguistics; corpus linguistics; cognitive mechanisms; mechanisms of speech generation.

В современном научном пространстве, характеризующимся междисциплинарностью научной мысли, речевой этикет выступает объектом разносторонних исследований, направленных на рассмотрение этикетных единиц через призму когнитивистики, прагматики, психолингвистики и других смежных направлений деятельности языкознания. Изучение англоязычного речевого этикета приобретает особую значимость в условиях глобализации и межкультурной коммуникации – современные лингвистические подходы стремятся выйти за рамки описания формул вежливости, обращая внимание на глубинные когнитивные механизмы, лежащие в основе речевого поведения [1, с. 20]. В этом контексте речевой этикет рассматривается не только как система устойчивых выражений, но и как динамичная структура, опирающаяся на универсальные и культурно-специфические концепты, такие как *politeness* ‘вежливость’ и *solidarity* ‘солидарность, единение’.

Стоит отметить, что англоязычный речевой этикет, который отличается определенной степенью опосредованной вежливости и обусловленностью культурными особенностями стран языка, является разветвленной системой и находит свое отражение в лексических и грамматических устойчивых конструкциях, а также принципах эвфемизации, косвенности и модальности. В этом случае когнитивная лингвистика, опирающаяся на представление о языке как отражении ментальных процессов, предлагает рассматривать речевой этикет не просто как совокупность правил речевого поведения, а как сложный феномен, репрезентирующий концептуализацию социальных ролей, моделей взаимодействия и культурных ценностей в сознании носителей языка [2].

Разнообразие средств выражения и стратегий современного речевого этикета требует все больше вариантов его продуктивного изучения для обеспечения поддержки социального равновесия и регулирования поведения участников речевого акта. В заданном контексте концепт вежливости стоит также рассматривать с точки зрения психолингвистики, а именно в качестве результата работы механизмов речевого восприятия, когнитивной обработки социально значимых коммуникативных паттернов и актуализации конкретных языковых единиц [3, с. 13].

Более того, представляется, что эффективный анализ конкретных ситуаций функционирования речевого этикета возможен благодаря активному развитию корпусной лингвистики и обогащению масштабных электронных корпусов английского языка – мощных инструментов анализа живого языкового материала в реальном контексте современного дискурса [4, с. 17]. Устойчивая эмпирическая основа позволяет объективно изучать продуцирование этикетных языковых конструкций через призму психолингвистических механизмов восприятия речи в естественном контексте, определяя частотность, специфику, контекст и прагматическую направленность использования формул вежливости.

Необходимо отметить, что с психолингвистической точки зрения речевой этикет представляет собой не просто набор устойчивых формул, а цель-

ную когнитивную структуру, в которую встроены фреймы и модели речевого поведения [5, с. 37]. Этикетные речевые единицы функционируют как прецедентные формы, активируемые в типичных коммуникативных ситуациях без необходимости синтаксической декомпозиции. Высокая степень «автоматизации» таких выражений обеспечивает быструю и социально адекватную реакцию, что особенно важно в условиях межличностного общения, требующего соблюдения норм вежливости и речевого такта.

Данная характеристика, отражающая автоматизацию и универсальность англоязычного этикета, делает такие формы продуктивными объектами корпусного анализа. Современные корпуса английского языка позволяют наблюдать за частотностью, контекстом употребления, жанровой спецификой и прагматической маркированностью этикетных конструкций в реальной коммуникации [4]. Например, корпусные данные могут подтвердить, что различные формулы приветствия, обращения, благодарности и др. преобладают в различных формах дискурса – как в устной речи, так и в письменных официальных текстах, а также позволяют получить конкретные статистические показатели такого употребления. Это не только иллюстрирует социодискурсивную обусловленность речевых стратегий, но и служит эмпирическим основанием для психолингвистических выводов о когнитивной доступности и речевой типичности таких выражений.

В заданном контексте корпусные исследования также могут послужить механизмом выявления **тенденций прагматической редукции**, а именно роста использования кратких формул в цифровом и интернет-дискурсе, отражающих адаптацию речевого этикета к новым форматам общения. Более того, совместное использование психолингвистических и корпусных методов исследования англоязычного речевого этикета позволит более продуктивно проводить **жанровую и стилистическую дифференциацию** (доминирование формул вежливого несогласия в СМИ, эмоционально окрашенных выражений в блогах и форумах), а также обозначать **гендерные и возрастные различия** в употреблении модальных, смягчающих конструкций или тенденцию к сокращенному речевому поведению [6].

Сопоставление полученных корпусных данных с психолингвистическими механизмами, такими как антиципация (предвосхищение), распознавание и регуляция эмоционального тона позволяет глубже понять, как носители языка выбирают и интерпретируют этикетные стратегии. Этикетные выражения, которые активируются спикером автоматически, одновременно функционируют как **когнитивные маркеры социального взаимодействия** и как естественные **психологические регуляторы** коммуникативного баланса, влияющие на восприятие статуса, дистанции и эмоционального фона общения.

Современные психолингвистические методы, включая айтрекинг и регистрацию вызванных потенциалов (ERP), предоставляют прямые данные о когнитивной доступности и скорости обработки речевых единиц. Так,

возможно предположить, что устойчивые этикетные конструкции вроде *Would you mind...?* ‘Вы не против...?’, *I’m afraid I can’t...* ‘Боюсь, что не смогу...’ запускают в мозге носителей языка автоматизированные речевые сценарии, что возможно подтвердить более быстрым временем фиксации взгляда и характерными ERP-реакциями. Это может свидетельствовать о высокой степени когнитивной закреплённости и минимальной нагрузке на рабочую память при их восприятии [3, с. 16; 7].

Однако в некоторых случаях психолингвистика может оказаться ограниченной в объеме и разнообразии речевых данных – более очевидное объединение этих подходов позволило бы сопоставить данные когнитивной обработки с реальной частотностью и контекстами употребления, например, выяснить, коррелирует ли высокая автоматизация восприятия формулы (по ERP-данным) с ее высокой частотностью и прагматической стабильностью в корпусе; идентифицировать зоны трудностей у изучающих английский язык, сравнивая их ERP-профили с типичными корпусными контекстами, а также создавать экспериментальные материалы на основе корпуса, отбирая типичные, нейтральные или, наоборот, редкие и контекстно обусловленные формулы для оценки когнитивной нагрузки и распознаваемости.

Следовательно, психолингвистика уточняет механизм, а корпус лингвистически «легитимизирует» материал. Их синтез позволяет выявить закономерности речевого этикета как когнитивной рутины, подкреплённой многократным языковым употреблением и дискурсивной релевантностью. Это открывает перспективы для создания более точных моделей речевого поведения и оптимизации методик преподавания английской прагматики.

Особое значение имеет изучение речевого этикета в межкультурной перспективе, так как психолингвистические исследования подтверждают, что при изучении английского языка как иностранного особую трудность представляет усвоение прагматических и социокультурных норм вежливости. В случае русскоязычных обучающихся наблюдается тенденция к переносу прямых стратегий родного языка, что может восприниматься в англоязычной культуре как избыточная прямолинейность или даже грубость [1, с. 21]. Таким образом, **интерференция когнитивных сценариев и стратегий вежливости** требует специального внимания при преподавании английского языка.

Корпусный анализ типичных ошибок, совершаемых изучающими английский, позволяет объективно зафиксировать влияние родного языка на выбор и интерпретацию речевых актов, недостаточную освоенность имплицитных норм речевой вежливости, а также различия в интерпретации прагматических маркеров речевого поведения. Интеграция психолингвистических и корпусных методов предоставляет мощную аналитическую базу для выявления закономерностей формирования прагматической компетенции у неанглоязычных обучающихся. На этой основе возможно создание более эффективных методик преподавания, ориентированных на интериоризацию речевых

стратегий, типичных для англоязычного коммуникативного пространства при учете контекста, целей общения и социальной роли собеседников [8, с. 41].

Более того, стоит также отметить, что анализ корпусных данных билингвальных и параллельных корпусов позволит изучать различия в реализации и восприятии вежливости у носителей разных вариантов английского языка или его культурных групп, указывая на наличие более или менее формального устойчивого выражения вежливости [9, с. 130–131].

Таким образом, в современной лингвистике наблюдается тенденция к интеграции эмпирических методов корпусного анализа и когнитивно-психологических моделей обработки речи. Особенно продуктивным представляется такой междисциплинарный синтез в изучении речевого этикета, который является важным компонентом прагматической компетенции и важным элементом когнитивной системы личности.

На стыке психолингвистики и корпусных исследований возможно формирование новой парадигмы анализа англоязычных этикетных выражений, позволяющей не только описывать, но и объяснять механизмы их восприятия, воспроизведения и вариативности в зависимости от социальных, культурных и коммуникативных условий.

Продуктивность анализа англоязычного речевого этикета на стыке психолингвистики и корпусных исследований обусловлена рядом предпосылок: **наличием достоверных эмпирических данных**, охватывающих широкий спектр речевых ситуаций, **возможностью выявления скрытых когнитивных и социокультурных закономерностей** использования этикетных выражений, а также **актуальностью задач межкультурной прагматики и высокой степенью автоматизации речевых формул**, что делает их удобными маркерами когнитивных и поведенческих паттернов.

Результатом такой междисциплинарной методологии может быть не только уточнение функционального статуса этикетных единиц, но и расширение теоретических моделей речевого поведения, интегрирующих данные лингвистики, психологии и социокультурной антропологии. В прикладном аспекте это способствует формированию устойчивых механизмов межкультурной компетентности, что делает исследование речевого этикета центральным элементом современного когнитивно-прагматического направления в лингвистике.

ЛИТЕРАТУРА

1. Галибердова Д. Р. Английский речевой этикет // Russian Journal of Education and Psychology. 2019. № 5. С. 19–24.
2. Соколова Н. Л. Английский речевой этикет: Монография / Н. Л. Соколова; Гос. ком. СССР по нар. образованию. М.: Изд-во Ун-та дружбы народов, 1991. 90, [2] с.

3. Федорова О. В. Психолингвистика vs. когнитивная лингвистика на карте современной когнитивной науки // Социо- и психолингвистические исследования. 2014. № 2. С. 7–20.
4. Захаров В. П. Богданова С. Ю. Корпусная лингвистика: учебник. 3-е изд., перераб. – СПб.: Изд-во С.-Петерб. ун-та, 2020. 234 с.
5. Вержбицкая Т. Н., Алексина Л. И. Психолингвистика: учеб. пособие. – Минск: БГПУ, 2009. 188 (205) с. URL: <http://elib.bspu.by/handle/doc/3238> (дата обращения: 25.06.2025).
6. Залевская А. А. Динамика общенаучных подходов к проблеме знания и некоторые задачи психолингвистических исследований // Вопросы психолингвистики. 2007. № 5. С. 1–8.
7. Culpeper, J. Conventionalised impoliteness formulae. *Journal of Pragmatics*. 2010. № 42(12). P. 3232–3245.
8. Комарова И. А., Коган М. С. Исследование английской фразеологии с помощью подходов корпусной лингвистики // Компьютерная лингвистика и вычислительные онтологии. 2019. Вып. 3. С. 40–49.
9. Кочетова Л. А. Лингвокультурные особенности в межкультурном деловом общении рассмотрение речевого этикета в деловой коммуникации английского и арабского языкового пространства // Дискурс. 2020. Т. 6, № 6. С. 126–136.

УДК 004.912:61:81'373.45

Евчик Надежда Семеновна
доктор филологических наук,
профессор кафедры речеведения
и теории коммуникации
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Nadejda Evchyk
Doctor of Philology,
Professor of the Department of Speech
Studies and Communication Theory
Belarusian State University
of Foreign Languages
Minsk, Belarus
nadejda.evchyk@gmail.com

Меньшенина Ирина Александровна
кандидат филологических наук,
старший преподаватель кафедры
иностранных языков
Витебский государственный
медицинский университет
г. Витебск, Беларусь

Irina Menshenina
PhD in Philology,
Senior Lecturer of the Department
of Foreign Languages
Vitebsk State Medical University
Vitebsk, Belarus
irinaslav23@rambler.ru

НАУЧНЫЙ ДИСКУРС И ТЕХНОЛОГИИ СМЫСЛОВОЙ ИНТЕРПРЕТАЦИИ В УСЛОВИЯХ ЦИФРОВОЙ ТРАНСФОРМАЦИИ: ОТ ЯЗЫКОВОГО БАРЬЕРА К ПРОФЕССИОНАЛЬНОМУ ВЗАИМОПОНИМАНИЮ

Статья посвящена анализу проблем восприятия англоязычных медицинских публикаций специалистами, для которых английский язык не является родным, и описанию возможностей их преодоления с помощью разработанной системы MedCompress. В ходе трехэтапного анкетирования были выявлены ключевые лингвистические барьеры: профессионально-детерминированный языковой инвентарь, спорадическое использование проформ, особенности авторского самовыражения, синтаксическая сложность англоязычных текстов. Для их преодоления создана интеллектуальная программа MedCompress, обеспечивающая автоматическую смысловую компрессию публикаций, генерацию рефератов и построение частотных словарей. Система поддерживает как классический, так и интерактивный режимы, что позволяет персонализировать результаты обработки под конкретные исследовательские задачи. Экспериментальное внедрение в медицинских университетах Беларуси показало рост продуктивности анализа публикаций более чем в четыре раза при сохранении качества понимания. Полученные результаты свидетельствуют о том, что цифровые технологии способны не только автоматизировать работу с текстами, но и формировать новое пространство профессионального взаимопонимания, снижая языковые барьеры и способствуя развитию научного диалога в условиях цифровой трансформации.

Ключевые слова: научный дискурс; профессиональная коммуникация; языковые барьеры; смысловая компрессия; автоматическое реферирование; система MedCompress.

SCIENTIFIC DISCOURSE AND TECHNOLOGIES OF SEMANTIC INTERPRETATION IN THE CONTEXT OF DIGITAL TRANSFORMATION: FROM LANGUAGE BARRIER TO PROFESSIONAL MUTUAL UNDERSTANDING

The article addresses the challenges faced by non-native English-speaking specialists in comprehending medical publications and demonstrates how these barriers can be overcome through the use of the MedCompress system. A three-stage survey identified key linguistic difficulties: professionally determined lexical inventory, sporadic use of preforms, specifics of authorial expression, syntactic complexity of English texts. To address these challenges, the MedCompress program was developed, providing automatic semantic compression of publications, generation of abstracts, and construction of frequency dictionaries. The system supports both classical and interactive modes, enabling personalized processing results tailored to specific research tasks. Experimental implementation in Belarusian medical universities demonstrated more than a fourfold increase in the efficiency of analyzing publications while maintaining comprehension quality. The findings confirm that digital technologies can not only automate text processing but also foster a new space of professional understanding, reducing linguistic barriers and supporting the development of scientific dialogue in the context of digital transformation.

Key words: scientific discourse; professional communication; language barriers; semantic compression; automatic summarization; MedCompress system.

Современные тенденции цифровой трансформации научной коммуникации все более четко очерчивают приоритеты смысловой интерпретации и компрессии знаний. Актуальность этого процесса особенно значима в области медицины, где объем англоязычных научных публикаций растет экспоненциально [1], а эффективное профессиональное взаимодействие между специалистами различных стран требует сокращения когнитивных и языковых барьеров [2, 3]. Эти барьеры препятствуют полноценному участию в международном научном диалоге и затрудняют использование последних достижений в клинической практике и исследовательской работе.

В научном пространстве ключевой проблемой становится не просто доступ к англоязычным публикациям, а возможность глубокого понимания их содержания специалистами, для которых английский не является родным языком. Профессиональные читатели – врачи и ученые – нуждаются в точном и полном осмыслении текста, сохраняющем логику, доказательность и новизну авторского замысла. В этом контексте задача интеллектуального анализа выходит за рамки традиционного реферирования, предполагающего обобщение и упрощение. Напротив, необходимо формировать условия для смысловой интерпретации, обеспечивающей сохранение глубины, последовательности и эпистемологической целостности научного изложения.

Специфика медицинской публикации заключается в ее прикладной направленности: она ориентирована на практическое решение задач здравоохранения, а потому предъявляет повышенные требования к уровню понимания текста со стороны специалиста. Врач, опирающийся в своей профессиональной деятельности на данные научных исследований, должен обладать не только актуальной базой теоретических знаний и клиническими навыками

ми, но и способностью к критическому мышлению, владеть основами доказательной медицины и уметь оценивать релевантность новой информации. Современный медицинский специалист обязан систематически знакомиться с последними достижениями в своей и смежных областях, самостоятельно искать и интерпретировать информацию по актуальным вопросам, объективно оценивать ее качество и применять в своей практике.

В условиях постоянно возрастающего объема публикаций, обработка англоязычного текста требует от читателя высокой когнитивной нагрузки, времени и специальных навыков. Как отмечается в методических рекомендациях по чтению научной медицинской литературы [4], понимание текста предполагает последовательную фиксацию ключевых идей, анализ незнакомых лексем, повторные прочтения и избирательное извлечение значимой информации, то есть по сути – ручную смысловую компрессию. Очевидно, что в реальных условиях лишь ограниченное количество публикаций может быть полноценно проанализировано медицинским специалистом с сохранением всех смысловых нюансов.

В то же время именно научные статьи обеспечивают динамическое распространение результатов новых исследований, способствуют передаче знаний и, следовательно, играют ключевую роль в развитии медицины и повышении качества медицинской помощи. Высокие требования, предъявляемые к жанру медицинской публикации, предполагают тщательное планирование исследования, сбор и верификацию данных, применение статистических методов и обеспечение внешней валидности результатов. Все это усиливает важность точного и глубокого понимания научного текста со стороны практикующего специалиста.

На этом фоне особенно актуальной становится задача выявления факторов, препятствующих адекватному восприятию англоязычных текстов. Проведенное в три этапа углубленное анкетирование 25 русскоязычных медицинских специалистов позволило квантифицировать основные лингвистические барьеры при чтении научных публикаций. Примененная прогрессивная методология, включавшая открытые вопросы, категориальный анализ и градуированную оценку, выявила четыре ключевые категории трудностей, каждая из которых характеризуется высоким уровнем сложности – от 82 % до 100 %.

Наивысший показатель сложности (100 %) зафиксирован в категории профессионально-детерминированного языкового инвентаря, отражающей многомерный терминологический вызов. Медицинская терминология, согласно лингвистическим исследованиям, насчитывает от 200 000 до 500 000 терминов, и является одной из наиболее обширных в научном дискурсе [5, с. 2]. Сложность усугубляется высокой степенью специализации: в современной медицине выделяется около 200 дисциплинарных направлений, каждое из которых развивает собственный лексикон. Дополнительные трудности создает система аббревиатур, отличающаяся как высокой степенью контекстуальной неоднозначности, так и полисемией. Особое место занимают эпонимы – собственные имена, используемые для обозначения болезней,

симптомов и методов лечения. Эта форма номинации требует от читателя знания истории медицины и культуры англоязычного медицинского сообщества, что нередко становится барьером интерпретации.

На втором уровне сложности (87 %) зафиксированы трудности, связанные со спорадическим использованием проформ и частыми лексическими повторами, которые, с точки зрения русскоязычных читателей, воспринимаются как стилистическая избыточность. Однако такая текстовая когезия объясняется структурными расхождениями между английской и русской языковыми системами: в англоязычном научном дискурсе повторы выполняют функцию поддержания логической связанности и акцентирования информации, тогда как в русскоязычной традиции предпочтение отдается разнообразию форм и средств выражения.

Следующая категория (82 %) связана с особенностями авторского самовыражения, в частности – допустимостью проявлений научной субъективности. Анализ анкет показал, что англоязычный академический стиль допускает элементы экспрессивности и аргументативного самопозиционирования, что вступает в противоречие с нормами нейтральности, характерными для русскоязычного научного письма. Это различие порождает трудности при интерпретации авторской позиции, особенно в случаях использования модальных конструкций и оценочных суждений, замаскированных под аналитическое изложение.

Наконец, еще одной ключевой проблемой оказалась синтаксическая сложность англоязычных текстов (98 %). Использование громоздких грамматических структур, включающих многоступенчатые подчинения, длинные конструкции и фразовую архитектуру с высокой степенью вложенности, затрудняет процесс семантического декодирования. Особенно сложными для восприятия являются субстантивные атрибутивные конструкции, характерные для научного стиля: последовательности существительных, выступающих в адъективной функции, часто не имеют прямых аналогов в русском языке. Такая языковая экономия и высокая информационная плотность требуют от читателя навыков сегментации и интерпретации смысловой иерархии компонентов высказывания.

Эмпирические данные, полученные в результате исследования, позволяют сделать важные выводы для развития технологий автоматической обработки научных медицинских текстов. Современные системы должны быть способны устранять неоднозначность терминов с полисемантической нагрузкой, адаптироваться к стилистическим и культурным особенностям дискурса, управлять синтаксической сложностью без потери семантической точности, а также сохранять эпистемологические и аргументативные нюансы оригинала. Преодоление этих трудностей требует не просто лингвистического перевода, а полноценного когнитивного и культурного посредничества. Все это обосновывает необходимость разработки интегрированных технологических решений, способных эффективно функционировать в условиях возрастающей сложности и глобализации научного медицинского дискурса.

Для решения поставленной задачи разработан программный комплекс MedCompress, который предназначен для работы с публикациями из трех ключевых областей современной медицины в международной практике: кардиологии, пульмонологии и гастроэнтерологии.

В основе архитектуры MedCompress лежит формализация процедуры генерации сжатой версии текста в форме реферата. Программа опирается на специально разработанное лингвистическое обеспечение, позволяющее выделять ключевые концепты и обеспечивать логически стройную редукцию содержания. Помимо собственно алгоритмов компрессии, в рамках проекта была разработана методика создания и ведения адаптивного специализированного медицинского словаря. Эта методика опирается на отбор лексических единиц из наиболее информативных фрагментов исходного текста, что делает словарь динамическим и способным к автоматическому расширению.

Функциональная реализация MedCompress организована в виде последовательности взаимосвязанных модулей, каждый из которых решает отдельную когнитивную задачу. В первую очередь, интерактивный модуль обеспечивает диалоговую загрузку текста и определение пользовательских параметров обработки. Пользователь может задать режим генерации реферата: классический – с автоматическим извлечением всех компонентов предметной суперструктуры или интерактивный – с выбором значимых для него разделов, таких как цели, методы или результаты. Такая возможность персонализировать выходной текст усиливает адаптивность программы к задачам конкретного специалиста.

Следующие модули обеспечивают предварительную обработку текста, структурный анализ, выделение ключевых концептов и генерацию реферата. Завершающий этап – построение алфавитно-частотного словаря на основе сгенерированного текста. Программа позволяет пользователю модифицировать словарь путем исключения так называемых "минус-слов" – второстепенных лексем, грамматических единиц, а также уже известных слов. Встроенный список стоп-слов включает артикли, предлоги, союзы и числительные. Кроме того, реализована возможность загрузки индивидуального списка исключений, что расширяет гибкость настройки результата.

MedCompress создает как полный, так и краткий частотный словарь. При этом краткий словарь исключает производные слова и сложные термины, которые могут быть поняты через знание составляющих элементов. Например, если в словаре присутствует глагол *to inflame* 'воспалить', то производные от него *inflammation* 'воспаление' и *inflammatory* 'воспалительный' могут быть исключены. Для лемматизации используется специализированная библиотека, приводящая все слова к начальной форме. Полученные словари пригодны как для учебных целей, так и для пополнения профессионального лексического минимума пользователя.

При обработке нескольких публикаций (до пяти) программа создает отдельные рефераты и словари для каждого файла, а также формирует сводные частотные словари. Эти словари сохраняются в тематически органи-

зованной базе, позволяющей накапливать и систематизировать лексические ресурсы. База может быть использована для сравнения словарей, выявления повторов и построения персонализированного корпуса терминов. Тем самым MedCompress выступает не только как средство автоматизированного реферирования, но и как инструмент создания лексико-семантического ядра научного дискурса в медицине.

В рамках пилотного внедрения программы в УО «Витебский государственный медицинский университет» и УО «Белорусский государственный медицинский университет» было обработано свыше 1000 англоязычных публикаций, как в классическом, так и в интерактивном режимах. Оценка качества программы осуществлялась методом анкетирования, в котором приняли участие 23 специалиста медицинского профиля. Анкета включала два блока: оценку удобства использования интерфейса и оценку соответствия сгенерированных рефератов заданным критериям. Полученные результаты показали высокий уровень удовлетворенности: 100 % респондентов положительно оценили функции перевода, скачивания и генерации лексического минимума; 87 % – удобство интерфейса и возможность сохранения данных. Наиболее значимыми критериями оказались точность передачи информации, доступность изложения, единый стиль и языковая четкость (100 %). За ними следовали логичность структуры (91,3 %) и полнота содержания (87 %).

Дополнительно был проведен эксперимент по сравнению эффективности анализа публикаций вручную и с использованием MedCompress. Две группы экспертов выполняли одну и ту же задачу – составление обзора литературы на основе 10 релевантных статей. Первая группа анализировала тексты вручную, затрачивая в среднем 28 минут на одну статью (2,14 статьи в час). Вторая группа использовала рефераты, сгенерированные программой: среднее время анализа составило 7 минут на статью, что соответствует 8,6 статьи в час. Таким образом, автоматизация процесса позволила повысить производительность труда более чем в четыре раза без утраты качества смысловой интерпретации.

Полученные результаты подтверждают высокую прикладную значимость интеллектуальной системы MedCompress как инструмента смысловой компрессии научных текстов. Ее функциональные возможности способствуют поддержке профессиональной коммуникации и развитию специализированной лексической базы, что особенно актуально в условиях стремительного развития цифровых технологий и трансформации научного дискурса. Важнейший вывод, вытекающий из проведенного исследования, заключается в том, что задачи, стоящие перед современной наукой, выходят за рамки простого перевода или механической обработки текста: они требуют интерпретации, способной учитывать как когнитивные, так и культурные особенности восприятия информации специалистами разных языковых и научных традиций.

Сформулированный на основе эмпирических данных подход демонстрирует, что вектор развития научной коммуникации все более уверенно сме-

щается от пассивного потребления информации к ее интеллектуальной переработке. Разработка и внедрение специализированных цифровых решений, таких как MedCompress, отражают становление новой парадигмы в научном дискурсе, в которой технологии выступают не как автономные инструменты, а как посредники между текстом и профессиональным пониманием.

В условиях экспоненциального роста англоязычных научных публикаций, особенно в таких критически важных сферах, как медицина, проблема языкового и когнитивного барьера приобретает системный характер. В этом контексте задача цифровизации научного знания заключается не столько в автоматизации чтения, сколько в создании механизмов, способствующих глубокой смысловой интерпретации, адаптированной к нуждам конкретного пользователя – специалиста, принимающего решения на основе анализа и синтеза сложной информации.

Практический опыт использования MedCompress показывает, что переход от традиционного линейного чтения к персонализированным формам смыслового восприятия текста позволяет существенно повысить эффективность междисциплинарной и межъязыковой коммуникации. Такие технологии открывают новые горизонты для профессионального взаимодействия, способствуют формированию единых понятийных оснований и обеспечивают более точную, оперативную и контекстно релевантную передачу научного знания.

Таким образом, результаты исследования позволяют утверждать, что научная коммуникация в условиях цифровой трансформации становится все более ориентированной на создание среды смыслового диалога. От языкового барьера мы переходим к формированию устойчивого пространства профессионального взаимопонимания – на основе современных технологий анализа, интерпретации и компрессии текста. Подобный сдвиг представляется ключевым для будущего международного научного сотрудничества, особенно в высокостандартизированных и социально значимых областях, таких как медицина и здравоохранение.

ЛИТЕРАТУРА

1. Johnson R. The STM report: an overview of scientific and scholarly publishing. 1968–2018. Celebrating the 50th anniversary of STM : fifth ed., Oct. 2018 [Electronic resource] // Oxford, UK : STM: International Association of Scientific, Technical and Medical Publishers, 2018. URL: <https://niascomm.in/wp-content/uploads/2019/03/STM-Report-2018.pdf> (date of access: 17.12.2024).
2. Торубарова И. И. Лингвокультурная специфика научного медицинского текста (на материале англоязыч. и русскоязыч. ст.) : дис. ...канд. филол. наук : 10.02.20. Тверь, 2021. 141 л.
3. Наumenко Ю. Н. Маркеры научных категорий исследования: сопоставительный анализ способов вербализации : дис. ...канд. филол. наук : 10.02.20. Тверь, 2021. 167 л.

4. Критический анализ медицинской литературы : метод. рекомендации / [Г. Р. Касян, А. В. Говоров, К. Б. Колонтарев, Д. Ю. Пушкарь] ; Правительство Москвы, Департамент здравоохранения города Москвы. – М. : АБВ-пресс, 2017. 29 с.
5. Кириллова Т. С. Пути формирования и лексикологические особенности английской терминологии подязыка медицины : (дерматология-венерология) : автореф. дис. ... канд. филол. наук : 10.02.04. Пятигорск, 1990. 17 л.

УДК 811.111' 42(045)

Зиневиц Надежда Васильевна

кандидат филологических наук,
доцент кафедры теории и практики
английского языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Nadzeya Zinevich

PhD in Philology, Associate Professor
of the Department of Theory
and Practice of the English Language
Belarusian State University
of Foreign Languages
Minsk, Belarus
nadzusha@gmail.com

Ермилина Анастасия Александровна

студент
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Anastasiya Yermilina

Undergraduate student
Belarusian State University
of Foreign Languages
Minsk, Belarus
anastasia.ermilina03@gmail.com

ЛИНГВОСТИЛИСТИЧЕСКИЙ АНАЛИЗ КАК ИНСТРУМЕНТ РАЗГРАНИЧЕНИЯ АВТОРСКИХ И РОБОТИЗИРОВАННЫХ НАУЧНЫХ СТАТЕЙ

Цель статьи – выявление лингвостилистических критериев для разграничения авторских и роботизированных научных статей, что связано с растущей адаптивностью современных нейросетей и их прогрессирующей способностью успешно имитировать тексты различной жанрово-дискурсивной принадлежности. Исследование проводится путем сопоставления диапазона и употребительности синтаксических выразительных средств, задействованных в научных статьях, созданных человеком и системами искусственного интеллекта. Результаты анализа показали, что обе группы текстов активно используют такие синтаксические ресурсы, как вводные слова и конструкции, повторы и параллелизм, что соответствует общей коммуникативной направленности научного дискурса. В то же время роботизированные научные статьи отличаются повышенной употребительностью таких прагматически нагруженных стилистических средств, как антитеза, вопросы в повествовании, градация и саспенс. Тем самым для сгенерированных с помощью ИИ научных текстов характерна повышенная экспрессивность и, как результат, тенденция размытия границ между научным и научно-популярным дискурсами. Таким образом, исследование доказывает эффективность использования лингвостилистического анализа как инструмента разграничения авторских и роботизированных научных статей.

Ключевые слова: экспрессивность; выразительные средства языка; синтаксические стилистические средства; научный дискурс; роботизированный текст.

LINGUO-STYLISTIC ANALYSIS AS A TOOL FOR DIFFERENTIATING HUMAN AND AI-GENERATED SCIENTIFIC ARTICLES

The aim of the article is to determine linguo-stylistic criteria for distinguishing human and AI-generated scientific articles in the current context of AI's growing adaptability and increasing

capacity to efficiently generate texts of different genres and discourses. The analysis is carried out by means of comparing the two groups of texts in terms of the range and frequency of syntactical expressive means used. The findings demonstrate that both human and machine-generated scientific articles make an active use of parentheses, parallelism and repetitions, which is justified by the main communicative goals of the scientific discourse in general. At the same time AI-generated texts differ in that they rely more on such pragmatically charged resources as antithesis, questions-in-the-narrative, climax and suspense. As a result the analysis reveals a greater expressiveness of automated scientific articles and their tendency to blur the boundaries between scientific-popular discourse and scientific discourse proper. The findings prove the efficiency of linguo-stylistic analysis as a tool for differentiating human and AI-generated scientific articles.

Key words: expressiveness; expressive means of language; syntactical stylistic devices; scientific discourse; AI-generated text.

Экспрессивность научного дискурса

Научный дискурс – характерный для науки способ организации речевой деятельности, который включает в себя специфические когнитивные (понятия, схемы, объекты, методы, парадигмы), социальные (запреты и предписания, приемлемые формы взаимодействия коммуникантов), а также, безусловно, лингвистические компоненты. Традиционно языковая специфика научного дискурса описывалась сквозь призму таких его ключевых характеристик, как обобщенный характер, объективность, точность, логичность и безличность изложения [1]. Как результат, практически полное отсутствие субъективности и экспрессивности постулировалось как одна из основополагающих черт научного дискурса в целом.

Вместе с тем современные исследования стремятся избегать столь категоричных утверждений относительно ограниченной экспрессивности в языке науки и не отказывают данной категории в праве на полноценное существование в данной сфере коммуникации. Так, было доказано, что субъективно отмеченные средства выразительности не только являются достаточно употребительными в научной коммуникации, но и выполняют здесь ряд важнейших прагматических функций. В частности, в научном дискурсе могут активно использоваться оценочные прилагательные (*outstanding, essential, important* и т. п.), интенсификаторы (*fully, completely, considerably* и т. п.), вводные слова и конструкции с семантикой личного мнения / степени уверенности автора (*apparently, clearly, perhaps* и т. п.) и разнообразные (преимущественно стертые) лексико-стилистические средства (например: метафора – *a driving force*, метонимия – *this paper shows...*, эпитеты – *elegant symmetry* и т. д.).

Однако в прагматическом плане важно, что в научном дискурсе данные средства будут употребляться не столько в своей «первичной» функции (усиление выразительности за счет создания эмоциональной окраски), сколько с целью:

- 1) более эффективного выражения мысли автора;
- 2) облегчения процесса восприятия сложной информации адресатом и увеличения прагматического эффекта сообщения;

3) повышения доказательной силы выдвигаемых аргументов [2, с. 99–100].

Тем самым, вопреки сложившемуся мнению, научный дискурс обладает экспрессивностью, однако она носит особый, «преобразованный», характер, поскольку в ее задачи в первую очередь входят более точная передача специализированной информации, эффективное закрепление знаний, а также обеспечение продуктивного взаимодействия между коммуникантами, что в конечном счете способствует прогрессу науки и общества в целом.

Также в контексте настоящего исследования особенно важно, что в языке науки при достаточно выраженной экспрессивной нагрузке лексических средств, наибольшую значимость все-таки приобретают ресурсы синтаксического уровня, в том числе синтактико-стилистические средства [2, с. 111]. Последние работы убедительно доказали, что наиболее частыми синтаксическими стилистическими приемами в научном дискурсе являются инверсия, обособление, вводные конструкции, параллелизм и повтор [2, с. 108]. Благодаря своей предсказуемости и структурированности, относительной простоте создания и широкому спектру возможных функций эти приемы органично вписываются в контекст академического повествования и активно участвуют в продвижении научного знания [3].

Как результат, изучение научного дискурса на современном этапе целесообразно проводить всесторонне, с учетом последних достижений лингвистической науки, принимая во внимание как особый экспрессивный потенциал научных текстов, так и наиболее активно задействованные для его создания языковые ресурсы.

Авторский и роботизированный научный дискурс

Современное научное знание развивается в условиях стремительной цифровизации, что сопровождается активным использованием технологий машинного обучения и искусственного интеллекта (ИИ) в процессах создания, анализа и распространения научной информации. Несмотря на отсутствие единого определения ИИ в научной литературе, под ним, как правило, понимаются модели, способные имитировать функции человеческого интеллекта – восприятие, обучаемость, логическое мышление, решение стандартных задач и даже творческую деятельность. Последние исследования уже доказали достаточно высокое качество роботизированных текстов различной жанрово-дискурсивной принадлежности, в том числе таких мало шаблонных жанров, как публичные выступления [4, с. 541].

В контексте настоящего исследования важно, что научные произведения отличаются достаточно высокой степенью клишированности и стандартизированности, что существенно облегчает их автоматическое генерирование. Следовательно, в сфере науки, где объективность, достоверность и точность сообщаемой информации имеют первостепенное значение, с появлением ИИ неизбежно возникают вопросы относительно авторства и степени личной и профессиональной ответственности создателей научных работ. В этой связи особый интерес представляет анализ языковых различий и сходств между

естественными (авторскими) и роботизированными (машинно-сгенерированными) научными текстами, что в конечном счете может способствовать их более успешному разграничению.

Целью данного исследования является сопоставительный анализ естественных и роботизированных научных статей с точки зрения их синтактико-стилистической организации. Данный ракурс исследования представляется перспективным, поскольку, как уже отмечалось выше, синтаксические стилистические ресурсы, с одной стороны, являются важным средством создания экспрессивности и, следовательно, повышения творческого потенциала научного произведения, что, как считается, является для ИИ более сложной задачей для имитации, нежели создание стандартизированных стереотипных текстов. С другой стороны, языковые средства синтаксического уровня по определению являются более предсказуемыми, в большей степени поддаются алгоритмизации и, по-видимому, могут легче генерироваться с помощью ИИ. Таким образом, изучение синтактико-стилистической организации естественных и роботизированных научных статей может позволить пролить свет на степень готовности современных ИИ-платформ продуцировать научные произведения во всей совокупности присущих им характеристик, что в конечном счете можно использовать для уточнения авторства научных работ.

Материалом для исследования послужили 10 авторских и 10 роботизированных научных статей. Естественные тексты были взяты из таких научных изданий, как HOW, Media and communication, Game, RUDN journal of language studies, semiotics and semantics, The journal of English teaching and applied linguistics, Vertimo studijos, Language learning and technology, The journal of specialized translation, Colombia applied linguistics journal. Роботизированные статьи были сгенерированы с помощью платформ ChatGPT, DeepSeek, Perplexity путем обращения к ним с запросами о создании текстов, сопоставимых с естественными по статусу автора (ученый-лингвист соответствующей научной квалификации), тематике (языковые особенности гейминг-коммуникации), объему (4–8 страниц) и жанровым характеристикам (научная статья).

Синтактико-стилистические особенности естественных научных статей

Процедура анализа включала в себя несколько этапов. На первом этапе отобранные тексты были подвергнуты лингвистическому анализу, в результате которого были выявлены используемые в данных текстах синтактико-стилистические приемы. На втором этапе проводился подсчет употребительности выявленных экспрессивных средств: устанавливалось, какой процент от общего числа обнаруженных синтаксических средств составляет конкретно взятый стилистический прием. На третьем этапе с опорой на контекст определялись прагматические функции экспрессивных средств в анализируемых жанрах. На завершающем этапе проводился сравнительный анализ, направленный на выявление сходств и различий языковой реализации синтактико-стилистических средств в естественных и роботизированных текстах.

Результаты анализа авторских научных статей приведены в таблице 1.

Таблица 1

Употребительность синтактико-стилистических средств
в авторских статьях

Стилистический прием	Употребительность, (%)
Вводные конструкции	20
Параллелизм	13,7
Повтор	13
Бессоюзие	13
Обособление	8,7
Инверсия	8,3
Антитеза	5,8
Многосоюзие	5,2
Эллипсис	3,8
Саспенс	3
Вопрос в повествовании	3
Градация	1,6

Рассмотрим подробнее полученные данные. Как видно из таблицы 1, наиболее употребительными синтактико-стилистическими средствами в проанализированных авторских текстах являются вводные конструкции, повтор, бессоюзие и параллелизм. Оценим результаты с точки зрения функциональной нагрузки выявленных средств в научном дискурсе.

Главной функцией вводных слов и конструкций является выделение и придание большей значимости дополнительным пояснительным и/или организующим текст замечаниям автора: *Therefore, when two professionals speaking different languages discuss a specialized field, a translator or interpreter is necessary...*; *In fact, even nowadays, more than three decades later, it is still rather difficult....* В данном примере вводные конструкции выполняют логико-смысловую функцию, направленную на логическую организацию текста и усиление аргументации. Вводное слово *therefore* указывает на причинно-следственную связь и подводит к выводу из предыдущего утверждения, тогда как *in fact* акцентирует внимание на значимости или неожиданности приводимого факта, усиливая убедительность высказывания [2, с. 109]. Иными словами, вводные элементы позволяют не только структурировать изложение, но и отразить позицию автора, внести дополнительные пояснения или уточнения, а также установить логические связи между фрагментами текста.

В свою очередь, основная функция повторов заключается в подчеркивании основных идей автора. Более того, повторение ключевых элементов также может быть использовано для привлечения к ним внимания зрителей: *The video game industry uses specialized terminology <...> The terminology is the most important pillar <...>*. В данном примере повтор выполняет функцию смыслового акцентирования. Повторение лексемы *terminology* подчеркивает ее значимость в контексте соответствующей статьи и помогает структурировать информацию, выделяя ключевую идею [2, с. 110]. Отсюда следует, что повтор служит мощным инструментом акцентирования, позволяющим выделить ключевые идеи и усилить ритмическую структуру научного высказывания.

Что касается *асиндетона*, то в научных статьях его используют для лаконичности и динамичности изложения, при этом данное средство помогает кратко перечислять факты, методы и результаты без лишних союзов, что соответствует строгому и нейтральному стилю академического текста: *Smartphones, tablets, Apple devices, browsers, and so on*. Функция асиндетона в данном случае заключается в придании изложению краткости и ритмической четкости. Отсутствие союзов ускоряет темп восприятия информации, делает перечисление более сжатым и выразительным, что особенно важно для научного стиля, стремящегося к точности, экономии языковых средств и структурной ясности [5, с. 226].

И, наконец, *параллелизм* обычно используется, когда хотят подчеркнуть сходство нескольких компонентов текста и одновременно уравнивать их значимость. В параллельных конструкциях делается попытка показать, что две или более идеи одинаково важны: *First step: vocabulary design: a. Determination of requirements <...>; b. Development of the macrostructure <...>; c. Development of the microstructure <...>*. Функция параллелизма заключается в структурировании текста и подчеркивании равнозначности или логической связанности нескольких элементов. В научном дискурсе параллельные конструкции способствуют ясности изложения, облегчают восприятие информации и акцентируют внимание на последовательности и/или системности излагаемых идей. Они помогают читателю быстрее сопоставлять и анализировать аналогичные части текста [2, с. 110].

Обобщая сказанное, подчеркнем, что активно используемые в естественных научных статьях синтактико-стилистические средства создают особую экспрессивность, которая в научной коммуникации понимается как экспрессивность акцентуации, конкретизации мыслей, логического подчеркивания, усиления аргументированной мысли автора, что в совокупности способствует, с одной стороны, более четкому выражению позиции автора, а с другой, активизации внимания читателя. Экспрессивные средства в научном тексте помогают выражению сложного для восприятия содержания, оптимизации общения в научной сфере и благодаря этому делают более убедительными и ясными доказательства, подчеркивают важные моменты высказывания, более точно выражают оценку автора описываемых явлений. Таким образом, ос-

новные функции экспрессии в научном дискурсе в первую очередь связаны с эффективной передачей научного знания и направлены на оптимизацию способа изложения [3; 6; 7].

Синтактико-стилистические особенности роботизированных научных статей

Проследим далее с помощью таблицы 2, как синтаксические стилистические средства реализуются в другом типе современных текстов – в научных статьях, сгенерированных искусственным интеллектом.

Т а б л и ц а 2

Употребительность синтактико-стилистических средств
в роботизированных статьях

Стилистический прием	Употребительность, (%)
Параллелизм	15,2
Повтор	13,6
Вводные конструкции	10,3
Антитеза	9,6
Вопрос в повествовании	9,3
Бессоюзие	7
Инверсия	6,7
Обособление	6,6
Градация	6
Многосоюзие	5,5
Эллипсис	5,2
Саспенс	5

Как видно из таблицы 2, и в авторских научных статьях, и в сгенерированных искусственным интеллектом текстах наблюдаются определенные сходства в использовании таких синтактико-стилистических средств, как параллелизм и повтор. Как отмечалось выше, эти средства выполняют важную функцию акцентуации ключевых идей, структурирования текста и повышения его доступности.

Однако, при более внимательном анализе становится очевидным, что в текстах, созданных с использованием ИИ-технологий, также достаточно активно используются такие «прагматически нагруженные» стилистические приемы, как антитеза, градация, вопросы в повествовании и саспенс. Рассмотрим эти средства более подробно.

Антитеза, как прием резкого противопоставления понятий или образов, используется для создания контраста и усиления выразительности текста. Этот прием позволяет ИИ подчеркнуть различия между идеями или явле-

ниями, делая текст более динамичным и запоминающимся: *Whereas films present a **predetermined path**, games offer a **multiplicity of choices**; The goal is not simply to inform, but **to immerse**; not merely to instruct, but **to engage**.* В примере функция антитезы заключается в создании контраста между противоположными идеями или явлениями с целью акцентирования различий между ними. В научном дискурсе антитеза помогает выделить ключевые противопоставления, усилить аргументацию и сделать выводы более четкими и выразительными. Такой прием способствует структурированию мысли, подчеркивая важные смысловые расхождения и усиливая риторическое воздействие текста [2, с. 110].

Градация, в свою очередь, представляет собой последовательное усиление эмоциональной нагрузки следующих друг за другом слов или фраз, что придает тексту эмоциональную насыщенность и драматизм, позволяя автору выстраивать риторически насыщенное и выразительное высказывание. Как результат, использование градации способствует удержанию внимания читателя: *Players do not merely consume content; they **interact** with it, **contribute** to it, **shape** it; <...> creating a dynamic narrative structure that **adapts, evolves, and responds** to the player's actions; <...> language becomes a tool not just for **communication**, but for **immersion**, for **identity construction**, for **narrative shaping**.* В данных примерах градация способствует привлечению внимания к ключевым моментам, подчеркивает развитие мысли и усиливает аргументацию, придавая тексту динамичность и стилистическую завершенность. Градация также помогает автору передать свое отношение к описываемым вещам и явлениям, выделить наиболее значимые или эмоционально насыщенные моменты, а также усилить впечатление от описания за счет динамического развития мысли [5, с. 219].

И, наконец, вопросы в повествовании применяются для более активного вовлечения читателя в размышления автора и создания иллюзии диалога. Этот прием помогает ИИ сделать текст более интерактивным и приближенным к человеческому общению: *With what latent energies do we blend planning and inference, formulating utterances that cascade into meaningful dialogue? Sequential language games, a concept introduced by Wittgenstein and enriched by subsequent scholars, compel us to recognize that language is less an artefact of rigid rules than a fluid performance of strategic improvisation and empathetic engagement.* Функция таких вопросов в научном дискурсе заключается в стимулировании когнитивной активности читателя и создании эффекта диалога между автором и аудиторией. Вопросы в повествовании направлены не на получение ответа, а на привлечение внимания к теме, формулировку ключевых проблем и побуждение к размышлению, что делает научный текст более живым, доступным и привлекательным для адресата [2, с. 110]. Кроме того, вопросы в повествовании могут служить средством перехода от известной информации к новой или использоваться для обозначения сложных, дискуссионных тем, не предполагающих однозначного

ответа. В таком случае они распределяются по тексту и побуждают читателя к размышлению, а в заключении могут выступать в качестве риторического акцента или предпосылки для дальнейшего научного исследования.

Таким образом, роботизированные научные статьи как проявляют определенные сходства со своими авторскими аналогами, так и отличаются от них в отношении использования более прагматически нагруженных стилистических средств. Подобная тенденция позволяет сделать важный вывод о наличии определенной стилистической специфики, характерной именно для автоматизированной генерации научных текстов. По всей видимости, ИИ, более активно встраивая в тексты элементы, типичные для публицистического и научно-популярного стиля, стремится сделать текст более «живым», «динамичным» и «убедительным» за счет повышенной экспрессивности и эмоциональной насыщенности. Это, вероятно, должно усилить привлекательность текста для широкой аудитории и способствовать активизации риторического воздействия.

Обобщая вышесказанное подчеркнем, что сопоставление синтактико-стилистических средств, применяемых в авторских и роботизированных научных статьях, позволяет выявить как общие черты, так и принципиальные отличия. Анализ показал, что наибольшее распространение среди синтаксических ресурсов, используемых в обеих группах текстов, имеют вводные конструкции, параллелизм и повтор. Эти средства, обладая высокой функциональной гибкостью, органично вписываются в структуру академического повествования, способствуя логической связности, ясности и выразительности текста, а в конечном счете установлению эффективного взаимодействия между участниками научной коммуникации.

В то же время искусственный интеллект активно использует более широкий спектр прагматически нагруженных стилистических приемов – таких как антитеза, градация, вопросы в повествовании – стремясь, по-видимому, к большей экспрессивности и приближению научного стиля к естественной человеческой речи, что приводит к смешению границ между научным и научно-популярным дискурсами.

Тем самым, в отличие от естественных текстов, где стилистические средства подбираются в соответствии с замыслом автора и всегда оправданы коммуникативной задачей, тексты, созданные ИИ, часто характеризуются шаблонностью и механической регулярностью использования выразительных конструкций. Это может приводить к избыточности или эмоциональной перегруженности, ослабляющей смысловую четкость. Как результат, в авторских статьях экспрессия, как правило, строго контролируется и служит логике аргументации, тогда как в ИИ-текстах она может быть применена без достаточного основания, что приводит к снижению научной точности и объективности. Отсюда следует, что авторские тексты демонстрируют более осознанный и контекстно мотивированный выбор выразительных средств, строго подчиненный целям аргументации и научной достоверности. Следовательно, несмотря на высокую способность искусственного интеллекта уже

сейчас достаточно эффективно имитировать академический стиль, на данный момент наблюдаются осязаемые отличия в глубине смысловой проработки, мотивации и характере использования выразительных средств.

Таким образом, настоящее исследование подтверждает, что несмотря на растущую адаптивность генеративных языковых моделей, между авторскими и роботизированными научными текстами сохраняются качественные различия. Это подчеркивает актуальность дальнейшего изучения стилистических особенностей ИИ-дискурса и необходимость развития инструментов для его идентификации. Подобные исследования важны как для лингвистической науки, так и для практики научной коммуникации в эпоху цифровых технологий, что особенно актуально в условиях активной цифровизации науки и образования.

ЛИТЕРАТУРА

1. Зорина Ю. В. К вопросу о некоторых специфических чертах научного дискурса [Электронный ресурс] // Международный научно-исследовательский журнал. 2018. № 8 (74). DOI: <https://doi.org/10.23670/IRJ.2018.74.8.027>.
2. Сопоставительная стилистика = Comparative stylistics / Т. П. Карпилович, П. П. Глазко, Т. Н. Гребень, Е. И. Конашенко, И. В. Метлушко, А. К. Шевцова. Минск : МГЛУ, 2016. 128 с.
3. Скрипак И. А. Синтаксические средства экспрессивности в текстах научного дискурса [Электронный ресурс] // Известия РГПУ им. А. И. Герцена. 2008. № 60. URL: <https://cyberleninka.ru/article/n/sintaksicheskie-sredstva-ekspressivnosti-v-tekstah-nauchnogo-diskursa> (дата обращения: 19.02.2025).
4. Зиневич Н. В. Стилистические средства выразительности как фактор разграничения авторских и роботизированных публичных выступлений // Язык в эпоху цифровых трансформаций и развития искусственного интеллекта : сб. науч. ст. по итогам междунар. науч. конф., Минск, 23–24 окт. 2024 г. Минск : МГЛУ, 2024. С. 536–542.
5. Гальперин И. Р. Стилистика английского языка. Москва : Издательская группа URSS, 2014. 336 с.
6. Арнольд И. В. Стилистика. Современный английский язык. Москва : Флинта/ Наука, 2002. 382 с.
7. Каштанова Д. В. Краткий анализ лингвостилистических особенностей научного дискурса английского языка [Электронный ресурс] // Проблемы современной науки и образования. 2018. № 3. Р. 50–52. URL: <https://cyberleninka.ru/article/n/kratkiy-analiz-lingvostilisticheskikh-osobennostey-nauchnogo-diskursa-angliyskogo-yazyka?ysclid=m3huanke2767988576> (дата обращения: 19.09.2024).

УДК 81 '32

Каличева Варвара Максимовна

студент департамента филологии

Научный исследовательский

университет «Высшая школа

экономики»

г. Санкт-Петербург, Россия

Varvara Kalicheva

Student of the Department of Philology

National Research University «Higher

School of Economics»

St. Petersburg, Russia

vmkalicheva@edu.hse.ru

КАРТОГРАФИЯ ЧУВСТВ В КИНОСЦЕНАРИИ: ПОСТРОЕНИЕ ЭМОЦИОНАЛЬНЫХ АРОК ПЕРСОНАЖЕЙ МЕЛОДРАМ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

В статье описываются результаты прикладного исследования, направленного на извлечение эмоционального содержания из диалогов главных героев мелодрам. Целью работы является проведение автоматической идентификации эмоционально насыщенных реплик с использованием разработанной методологической основы для применения инструментов анализа тональности к диалогической речи. Для этого предлагается метод, интегрирующий подходы компьютерной лингвистики и теорию базовых эмоций, который позволяет строить эмоциональные арки (*emotional arcs*) персонажей. В основе метода лежит последовательная классификация реплик с использованием модели DistilRoBERTa-base, настроенной на распознавание шести базовых эмоций, с последующей агрегацией результатов для визуализации арок. Автоматизация процесса обеспечивается специально разработанным скриптом на Python, выполняющим задачи извлечения диалогов из сценариев, предобработки текста, пакетной обработки моделью и построения арок. Материалом исследования послужил корпус диалогов главных героев из десяти англоязычных фильмов в жанре мелодрамы, выпущенных в период с 1990-х по 2010-е годы. Общий объем — 12 749 реплик (67 841 словоупотребление). Практическая значимость работы заключается в возможности применения автоматически построенных эмоциональных арок для обогащения метаданных фильмов в рекомендательных системах на основе паттернов эмоционального развития персонажей, а также анализа связи между динамикой эмоций в нарративе и зрительским восприятием (*consumer sentiment tracking*).

Ключевые слова: анализ тональности; эмоция; фильмы; эмоциональные арки; машинный метод.

MAPPING OF EMOTIONS IN FILM SCREENPLAYS: CONSTRUCTING EMOTIONAL ARCS OF MELODRAMA CHARACTERS USING MACHINE LEARNING METHODS

This article presents the results of an applied study focused on extracting emotional content from dialogues of main characters in melodramas. The work's objective is to conduct automated identification of emotionally charged utterances using a developed methodological framework for applying sentiment analysis tools to dialogic speech. To achieve this, we propose a method integrating computational linguistics approaches and the basic emotion theory, enabling the construction of characters' emotional arcs. The method's foundation involves sequential utterance classification employing a DistilRoBERTa-base model configured for recognizing six basic emotions, followed by result aggregation for arc visualization. Process automation is implemented through a custom Python script that handles dialogue extraction from scripts, text

preprocessing, batch processing via the model, and arc construction. The research material consists of a corpus of main character dialogues from ten English-language melodrama films released between the 1990s and 2010s, totaling 12,749 utterances (67,841 words). The practical significance lies in applying automatically generated emotional arcs for enriching film metadata in recommendation systems based on emotional character development patterns, and for analyzing correlations between narrative emotion dynamics and audience perception (consumer sentiment tracking).

Key words: sentiment analysis; emotion; films; emotional arcs; machine learning approach.

Анализ тональности (англ. *sentiment analysis*) представляет собой метод обработки естественного языка для обнаружения и извлечения субъективной информации, такой как мнение, отношение и эмоция, из языковых данных [1, р. 1–2]. Сложность, связанная с определением эмоциональной окраски текста, варьируется в зависимости от применяемого подхода. Верхнеуровневая классификация основана на бинарной системе, при которой тексту присваивается один из полярных эмоциональных лейблов – позитивный или негативный; в некоторых моделях также вводится дополнительная категория – нейтральный [2, р. 110694]. В ряде методов анализа тональности текстам присваиваются оценки по заранее заданной шкале, например: от –2 до +2. Подобный подход, как правило, ассоциируется с понятием сентимента, представляющего собой обобщенную аффективную реакцию.

Анализ на более глубоком уровне включает определение не просто валентности высказывания, а его эмоциональной окраски с опорой на психологические модели эмоций. История изучения эмоций имеет богатую традицию, восходящую к работам XIX века, в частности – к публикации книги Ч. Дарвина «Выражение эмоций у человека и животных» (1872) [3]. В современном контексте, особенно в области компьютерного анализа эмоций, наибольшее распространение получили три основные теории: теория базовых эмоций Экмана (*theory of basic emotions*) [4], колесо эмоций Плутчика (*wheel of emotion*) [5], циркумplexная модель (*circumplex model*) [6].

В этом контексте в терминологическое поле анализа вводится понятие эмоции. Таким образом, возникает необходимость четкого разграничения понятий сентимент и эмоция, отражающих различные уровни аффективной интерпретации текста. В статье [7] авторы указывают, что сентимент представляет собой позитивное или негативное чувство, лежащее в основе мнения. Однако дать универсальное определение для понятия «эмоция» представляется затруднительным: отмечают разногласия в методологическом аспекте вследствие наличия множества входящих в понятие конструкторов (экспрессия, поведение, физиология и феноменология) [Там же]. Тем не менее, с опорой на работу [8, р. 508], эмоция может быть охарактеризована как совокупное состояние чувств, включающее в себя физиологические изменения, отражающиеся в готовности к действию, когнитивное восприятие ситуации и внутренние переживания, возникающие в результате оценки себя или обстоятельств.

Что касается объектов исследований в области анализа тональности, то ими, как правило, выступают компании, их продукты и услуги, о которых публиковались отзывы в интернете, что широко используется в бизнес-аналитике [9] и маркетинге [10] для выбора стратегии взаимодействия с покупателями [11]. Помимо этого, метод применяется и разрабатывается в других областях: в экономике для прогнозирования изменений на рынке ценных бумаг [12], в медицине для выявления и отслеживания эмоционального состояния пациентов [13], в политологии для прогнозирования результатов выборов [14].

В менее практико-ориентированной сфере применения сентимент-анализа решаются в основном филологические задачи на материале художественных произведений – например, с целью идентификации художественных жанров [15], составления топографических карт эмоций (например, викторианского Лондона [16]), изучения корреляции между работами автора и его репутации [17], определения динамики настроений между персонажами [18] и выявления эмоциональной арок, определяющих универсальные эмоциональные траектории (сменяющие друг друга падения и подъемы) книг [19].

В последнее время исследования анализа тональности активно смещаются в сторону кинематографического материала под влиянием трех взаимосвязанных факторов: прогресса мультимодальных технологий, таких как архитектура MulT [20], устранившего барьеры для изучения кино, интегрируя обработку текста (NLP), визуала (CV) и звука; коммерческого запроса индустрии на изучение связи между эмоциональным содержанием диалогов и зрительской вовлеченностью для прогнозирования кассовых сборов; уникальности кинодискурса как структурированного источника для анализа протяженных нарративов и культурных кодов.

Множество подходов к анализу текстового материала с использованием инструментов сентимент-анализа применяются для решения как теоретических, так и прикладных задач. Эти подходы традиционно классифицируются на две основные группы методов: лингвистические и основанные на машинном обучении [21].

Метод на основе правил (*rule-based*) использует композиционную семантику для определения эмоциональной окраски слов в локальном контексте [22]. Процесс включает применение логистических правил («если → то») [23]; поэтапную проверку токенов – при наличии лексемы в лексиконе (ручной/автоматической генерации) ей присваивается тональный лейбл (напр., «+1» для позитивной тональности) [24] и суммирование меток для оценки тональности предложения с учетом контекстуального влияния слов [25]. Основные преимущества метода – высокая интерпретируемость результатов и точность (при качественных правилах). Главный недостаток – значительные временные/финансовые затраты на создание лексиконов, особенно для языков с ограниченными ресурсами, например, русского [2]. Метод на

основе словарей (*dictionary-based methods*) прост в применении, поскольку не требует обучающие данные и для классификации. Достаточно иметь текстовую коллекцию, а именно вручную или автоматически размеченный словарь с эмоциональной лексикой (*affective lexicon*), в котором будут представлены списки позитивных и негативных слов, то есть каждому токену будет присвоена оценка на основе полярности [21]. Корпусный метод анализа тональности подразумевает создание уникального корпуса данных, адаптированного для конкретной цели или предметной области и включающего различные типы текстов: от отзывов до литературных произведений, что позволяет более точно присвоить лейбл в зависимости от контекста [26].

Машинное обучение, применяемое в задачах анализа тональности текста, может быть реализовано в двух основных парадигмах – с учителем (*supervised learning*) и без учителя (*unsupervised learning*). Обучение с учителем (логистическая регрессия [27], метод опорных векторов [28]) обеспечивает высокую точность при наличии качественной разметки, но требует значительных ресурсов для аннотирования данных. Обучение без учителя (кластеризация) работает по принципу автоматического выявления скрытых закономерностей в текстах, однако уступает в точности и интерпретируемости, требуя больших объемов данных для устойчивых эмбедингов. Глубинное обучение (трансформеры BERT [29], векторное вложение слов GloVe [30]) обрабатывает сложные структуры и большие данные, обеспечивая высокую точность и поддерживая перенос обучения (*transfer learning* [31]), но сохраняет зависимость от вычислительных ресурсов и контекста [32]. Гибридные подходы (например, интеграция словарей с методами опорных векторов [33]) компенсируют ограничения отдельных методов, комбинируя лингвистические правила с машинными, однако наследуют их сложности.

Большинство исследований, посвященных анализу настроений на материале фильмов, сосредоточено на извлечении имплицитно выраженной эмоциональной информации в тексте рецензий. Например, исследователи К. Lu и J. Wu проводят подобный анализ с целью облегчить деятельность сотрудников медиаплатформ, которые заинтересованы в понимании предпочтений аудитории в отношении фильмов [34]. Исследования в этой области едины в своей цели (создание привлекательного для потребителей контента на основе их реакций), однако проявляют методологическое разнообразие в подходах: от машинного обучения [35] до гибридных моделей [36] и лексико-ориентированного анализа [37]. Особый интерес представляет мультимодальный анализ (аудиовизуальный сентимент-анализ) [38], где сверточные нейросети, обученные на визуальных и аудиоданных голливудских фильмов, строят эмоциональные арки, отражающие ключевые аспекты сюжетов и вовлеченность аудитории.

В последние годы наблюдается хоть и небольшая, но растущая тенденция к использованию методов сентимент-анализа в изучении нарративной структуры кинематографических произведений. Академический

интерес к исследованию нарративной структуры фильмов с точки зрения эмоционального обусловлен, во-первых, тем, что выбор сценария является одним из основополагающих решений в процессе кинопроизводства. Во-вторых, эмоции рассматриваются как ключевой элемент конструирования реальности, обеспечивающий непрерывное поддержание внимания аудитории [39, p. 28].

Исследования направлены на отслеживание эмоциональной динамики раскрытия образа персонажей через вычисления эмоциональной окраски использованных слов [40, p. 7–8] и разделяются на два подхода: усреднение по всему сценарию (с сохранением ремарок), примером которого служит выявление шести универсальных эмоциональных арок через анализ 6 174 сценариев для оптимизации генерации контента в индустрии развлечений [41], и детальный анализ диалогов, где фокус на ключевых героях позволяет моделировать индивидуальные эмоциональные траектории – как в работе [40], впервые применивших UED-метрику для измерения скорости и интенсивности отклонений эмоций от базового состояния.

В настоящем исследовании рассматриваются эмоциональные арки диалогов главных героев фильмов с выраженной любовной линией, вышедшие в период с 1990 по 2010. Гипотеза состоит в том, что в речи персонажей отражаются характерные для жанра эмоциональные паттерны. Вследствие того, что жанр рассматривается как фактор, предопределяющий доминирующую эмоцию, испытываемую зрителем, для ее проверки был выбран жанр мелодрамы. Отражая социально-политические трансформации и вовлекая аудиторию через апелляцию к повседневному опыту [42], мелодрама способна вызывать эмпатию в условиях культурного и социального разнообразия [43]. Она синтезирует личностные и общественные уровни опыта, формируя прочную эмоциональную связь с аудиторией [44]. Именно эта способность к эмоциональному воздействию является релевантным объектом академического анализа в контексте настоящей работы. Выбранная периодизация обусловлена значительными трансформациями жанра, произошедшими в период с 1990-х по 2010-е годы: рост популярности жанра, так и усложнение его эмоциональной и нарративной структуры, что находит эмпирическое подтверждение в исследовании [45], где на примере сцены «известие о смерти» из фильма «21 грамм» (2003) демонстрируется, что мелодраматический дискурс способен инициировать интенсивную эмоциональную траекторию, охватывающую более десяти эмоций.

Материалом послужили диалоги главных героев (мужчины и женщины) из десяти англоязычных мелодрам 1990–2010-х гг., получивших пользовательскую оценку не ниже 7 из 10 возможных баллов: «Титаник» (*Titanic*), «Красотка» (*Pretty Woman*), «Покидая Лас-Вегас» (*Leaving Las Vegas*), «Лучше не бывает» (*As Good as It Gets*), «Жестокие игры» (*Cruel Intentions*), «Осень в Нью-Йорке» (*Autumn in New York*), «Холодная гора» (*Cold Mountain*), «Еще одна из рода Болейн» (*The Other Boleyn Girl*), «500 дней

лета» (*500 Days of Summer*), «Трудности перевода» (*Lost in Translation*). Все тексты были взяты с платформы The Internet Movie Script Database (IMSDb) – одного из крупнейших англоязычных онлайн-ресурсов, специализирующихся на публикации оригинальных киносценариев. Выбор англоязычных текстов обусловлен недостатком качественных, масштабных и общедоступных коллекций размеченных данных на русском языке, необходимых для обучения и оценки моделей машинного обучения в задачах анализа тональности [46].

В связи со спецификой поставленной исследовательской задачи – создание эмоционально-речевых портретов персонажей на основе их диалогов из киносценариев – текстовый материал прошел этап предварительной экспертной обработкой, а именно были исключены ремарки, описания сцен, плана съемки, а также технические указания режиссеру и оператору. Далее текст каждого сценария был структурирован в табличный формат при помощи разработанного кода на языке Python, где в первом столбце автоматически указывалось имя персонажа, а во втором – соответствующая ему реплика.

В результате обработки были выделены тексты, состоящие исключительно из реплик. Количество реплик и общее число слов в них приведены в Таблице.

Распределение количества реплик и слов по фильмам

Название фильма	Количество реплик	Количество слов
Титаник (Titanic)	2090	11 614
Красотка (Pretty Woman)	2753	12 985
Покидая Лас-Вегас (Leaving Las Vegas)	1223	7459
Лучше не бывает (As Good as It Gets)	1835	10 184
Жестокие игры (Cruel Intentions)	1790	9670
Осень в Нью-Йорке (Autumn in New York)	1714	10 102
Холодная гора (Cold Mountain)	1663	10 145
Еще одна из рода Болейн (The Other Boleyn Girl)	1818	8581
500 дней лета (500 Days of Summer)	1923	8084
Трудности перевода (Lost in Translation)	783	3616
Всего	12 749	67 841

Эмоциональный анализ проведен с применением модели DistilRoBERTa-base – трансформера для классификации эмоций в текстовых данных на

английском языке, которые не являются явно эмоциональными [47]. Основопологающим фактором выбора стало ее дообучение на транскриптах сериала «Друзья», структурно и функционально близких к кинодиалогам. Дополнительным достоинством модели является ее способность предсказывать все шесть эмоций по общепризнанной в академическом поле [48; 49; 50] классификации Экмана и нейтральный класс. Такое широкое применение теории эмоций обусловлено, с одной стороны, универсальность модели, а с другой – ее прикладной эффективностью в распознавании выраженных посредством текста эмоциональных состояний при обработке естественного языка. Кроме того, согласно работе [40, р. 3], именно базовые эмоции, такие как гнев, радость, страх и грусть, являются наиболее частотными при вербализации в повседневной коммуникации.

В результате применения модели анализа эмоциональной тональности были получены линейные графики (см. Рисунки 1–10). В рамках анализа эмоционального содержания широко применяется концепция эмоциональных арок, которые, не предоставляя прямой информации о сюжете, являются частью повествовательной структуры, соотносимые с развитием мотивов. Эмоциональные арки, визуализированные через линейные графики, отражают динамику настроений в нарративной структуре, соотносясь с развитием мотивов [51, р. 1–2]. Текст сегментирован на последовательные окна (от абзаца до эпизода), где ось X представляет нормализованный хронометраж, а ось Y – интенсивность аффективных меток. Размер окна определяет детализацию: меньшие интервалы выявляют резкие изменения (риск шума), большие – сглаживают колебания.

Ключевыми особенностями применения модели к тексту включают следующие аспекты.

1. Процентное распределение по тексту

Ось абсцисс графиков (ось X) соответствует процентному соотношению от общего количества реплик в диалогах, что позволяет отслеживать изменения эмоциональной окраски по мере продвижения сюжета – от 0 % (начало) до 100 % (конец реплик).

2. Гауссово сглаживание

Для устранения резких колебаний в значениях интенсивности эмоций и более четкого выявления трендов используется Гауссов фильтр с параметром, равным отношению общего количества строк к 50 ($\text{total\_rows} / 50$). Такой подход позволяет получить сглаженные кривые, сохраняющие основные закономерности при подавлении шумовых выбросов.

3. Визуализация нескольких эмоций одновременно

На одном графике одновременно отображаются кривые для шести базовых эмоций (радость, печаль, страх, отвращение, гнев, удивление), каждая из которых представлена отдельным цветом из цветовой карты *rainbow*, что обеспечивает наглядность графика.

4. Нормализация данных

Масштабирование графика осуществляется по оси ординат (ось Y) в интервале от 0 до 1, а по оси абсцисс от 0 % до 100 % с шагом в 10 %, что облегчает сопоставление интенсивности эмоций между собой в различные моменты текста.

5. Исключение нейтральной тональности

Нейтральные значения были исключены из анализа, так как в рамках настоящего исследования они не представляют существенной ценности для последующего анализа, поскольку основное внимание сосредоточено на выраженных эмоциональных реакциях.

6. Подсчет и вывод соответствия строк процентам

Функция реализует сопоставление между процентами и конкретными строками текста, что необходимо для ручного анализа, применяемого в настоящей работе, так как позволяет точно определить, какая строка сценария соответствует определенному проценту прохождения текста.

Общий анализ эмоциональной интенсивности, позволяет выявить ряд тенденций, а именно преобладающие на протяжении всех кинофильмов эмоции и моменты наибольшей эмоциональной активности. Рассматриваемые эмоции получают выражение на лексическом уровне, поскольку методы анализа тональности при извлечении эмоциональной окраски содержания текста, опираются именно на него.

Среди шести эмоциональных состояний наибольшую интенсивность демонстрируют *удивление* (*surprise*), *радость* (*joy*) и *печаль* (*sadness*), значения которых варьируются в диапазоне от 0.08 до 4.43 (см. Рисунок 1–10). Высокая частотность выражения эмоции *удивления* обусловлена особенностями нарративной структуры мелодраматических произведений, в которых интрига нередко строится на сюжетных поворотах, противоречащих зрительским ожиданиям [52, с. 65]. Преобладание позитивно окрашенной лексики в диалогах персонажей, независимо от жанровой принадлежности фильма, подтверждено эмпирическими данными: в ходе анализа корпуса включающего более тысячи киносценариев установлено, что доля лексем, выражающих доверие, надежду (или ожидание) и *радость*, составляет 20,3 % от общего числа эмоционально маркированных слов [40, р. 7–8]. Это свидетельствует о доминировании оптимистичных реплик в кинодискурсе в целом. Тем не менее, в исследовании фильма «Стелла Даллас» (1937), посвященном изучению именно мелодраматического нарратива, было установлено, что центральной эмоцией жанра является грусть [53, р. 30–37]. Таким образом, выявленные в настоящем анализе диалогических реплик доминирующие эмоции удивление, радость и *грусть* коррелируют с существующими теоретическими и эмпирическими исследованиями в области кинематографа.

Следует отметить также, что эмоциональные арки гнева в некоторых случаях достигают значений, которые могут быть сопоставимы с эмоциональными арками доминирующих эмоций (*радости, удивления, печали*),

однако остаются в пределах значения 0.27 по оси Oy (см. Рисунок 1–10). Вероятно, это связано с проявлением гнева в контексте напряженных ситуаций и драматических поворотов, на фоне которых развивается сюжет и формируются поведенческие стратегии персонажей в рамках романтических отношений. Кроме того, в структуре мелодраматического жанра особое место нередко отводится образу героини-жертвы, чья судьба сопряжена с множеством страданий и негативных событий, что может способствовать проявлению гнева как ключевой эмоции в определенном фрагменте фильма.

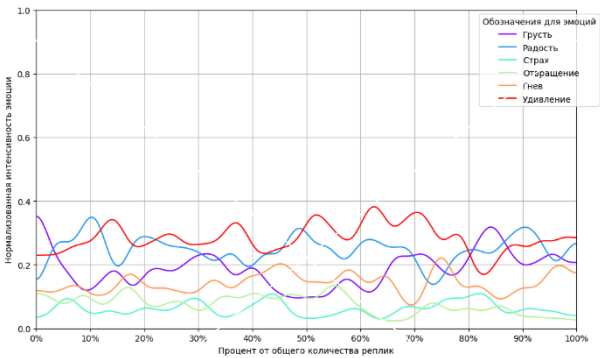


Рис. 1. Динамика тональности диалогической речи персонажей фильма «Осень в Нью-Йорке»

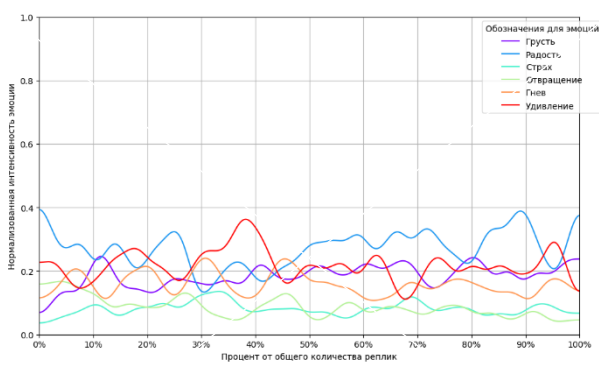


Рис. 2. Динамика тональности диалогической речи персонажей фильма «Лучше не бывает»

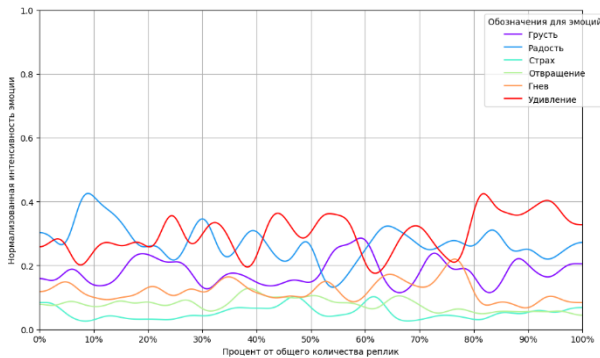


Рис. 3. Динамика тональности диалогической речи персонажей фильма «500 дней лета»

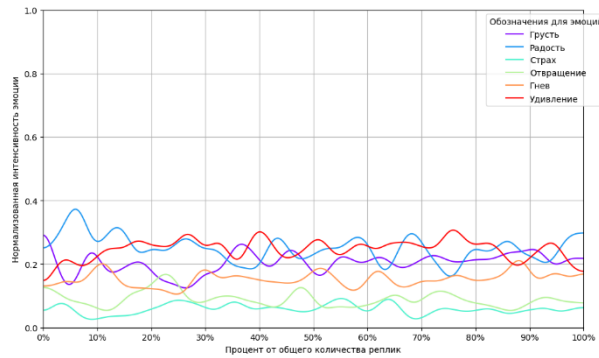


Рис. 4. Динамика тональности диалогической речи персонажей фильма «Жестокие игры»

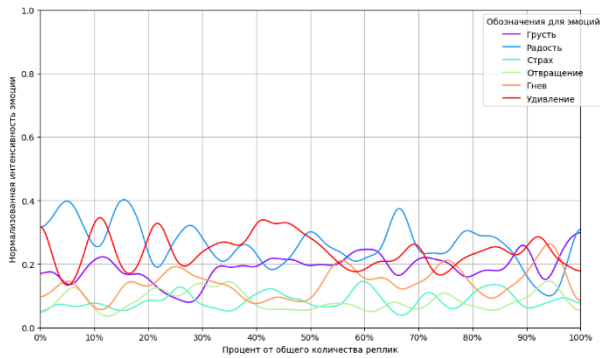


Рис. 5. Динамика тональности диалогической речи персонажей фильма «Покидая Лас-Вегас»

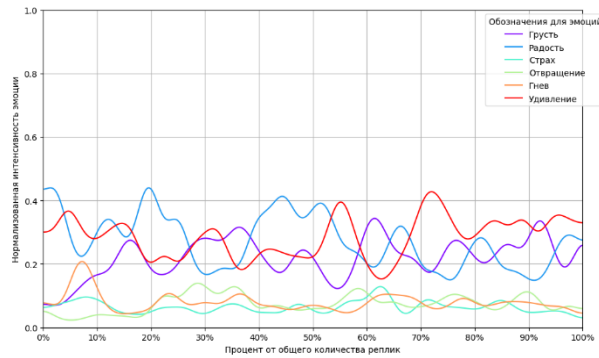


Рис. 6. Динамика тональности диалогической речи персонажей фильма «Трудности перевода»

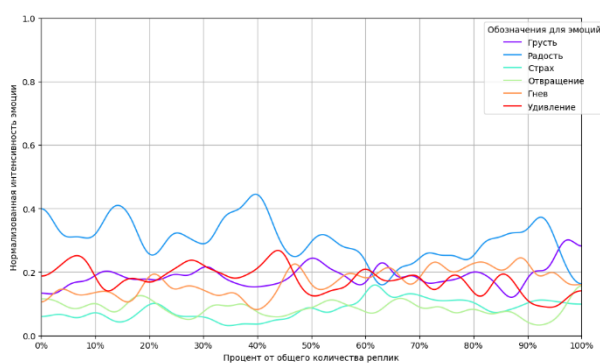


Рис. 7. Динамика тональности диалогической речи персонажей фильма «Титаник»

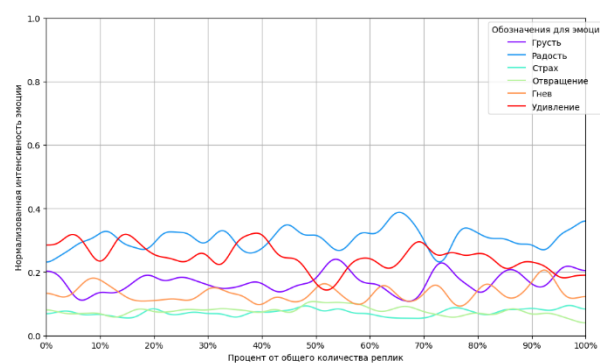


Рис. 8. Динамика тональности диалогической речи персонажей фильма «Красотка»

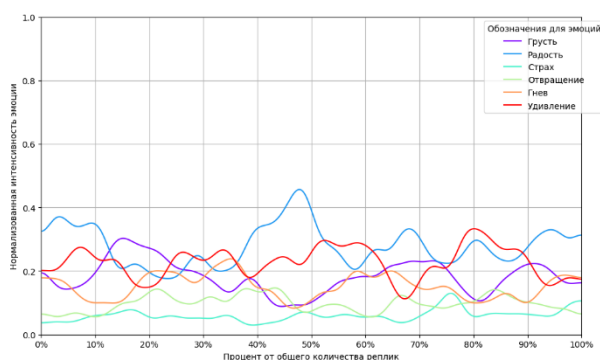


Рис. 9. Динамика тональности диалогической речи персонажей фильма «Холодная гора»

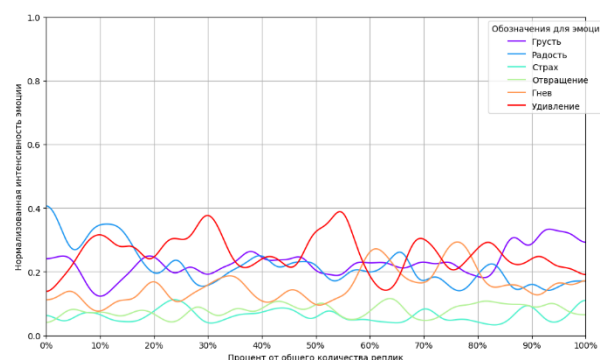


Рис. 10. Динамика тональности диалогической речи персонажей фильма «Еще одна из рода Болейн»

Обращаясь к анализу участков наибольшей эмоциональной интенсивности, необходимо отметить, что установить единый паттерн эмоциональной траектории для всех фильмов не представляется возможным. Тем не менее, в большинстве случаев эмоциональные пики фиксируются в интервале от 40 до 50 % общей продолжительности сценарного текста, что скорее всего связано с развитием основного конфликта или с неожиданными сюжетными поворотами, способствующими резкому усилению эмоциональной насыщенности диалогов персонажей. Так, в «Титанике» эмоциональный пик по эмоции *радость* приходится на 40 % процентов фильма – момент, когда Джек внезапно попадает на светский ужин, на которой он поднялся из третьего класса, оказав услугу Роуз.

Однако в ряде фильмов, таких как «Осень в Нью-Йорке» (Рисунок 1), «Лучше не бывает» (Рисунок 2) и «500 дней лета» (Рисунок 3), пиковые эмоциональные моменты смещаются к финалу повествования и приходятся на диапазон от 70 до 90 % от общего объема текста. Это согласуется с нарративной логикой классической драматургической структуры, согласно которой кульминация происходит ближе к концу истории – ориентировочно на 90 % ее длительности [3], что позволяет достичь наибольшего эмоциональ-

ного отклика и обеспечить завершенность сюжетной арки. Вероятно, это находит отражение в характере и эмоциональной интенсивности речевых взаимодействий персонажей.

В фильмах «Жестокие игры» (Рисунок 4) и «Покидая Лас-Вегас» (Рисунок 5) пиковые моменты выражены менее отчетливо, что может быть связано с более равномерным распределением эмоционального напряжения. В «Трудностях перевода» (Рисунок 6) и «Титанике» (Рисунок 7) эмоциональная динамика охватывает широкий отрезок от 20 до 80 % нарративной структуры, демонстрируя высокий уровень напряженности на протяжении значительной части повествования. В «Красотке» (Рисунок 8) наблюдаются два выраженных эмоциональных пика, приходящихся на примерно 40 % и около 65 % от общего объема текста.

Таким образом, эмпирически подтверждено, что речь персонажей мелодрам отражает характерные жанровые паттерны, проявляющиеся в доминировании трех эмоций – удивления, радости и печали, что коррелирует с нарративной спецификой жанра: сюжетными поворотами, канонической ролью грусти и универсальностью позитивной лексики в кинодиалогах. Что касается распределения эмоциональных пиков, здесь наблюдается вариативность: лишь 30 % фильмов соответствуют классической драматургической модели, при которой кульминация приходится на 90 % хронометража, тогда как остальные демонстрируют индивидуальные траектории. Эта гибкость не отменяет общих механизмов, но указывает на влияние конкретного сюжетного конструктора. Таким образом, гипотеза о связи диалогов с жанровыми константами верна для базовых эмоций, но требует дальнейшего изучения контекстно-обусловленных отклонений.

Исследование формирует первичную научно-методологическую базу для дальнейших комплексных работ по анализу эмоционального содержания текста в отечественном научном сообществе и вносит вклад в развитие современных подходов к изучению кинодискурса в языковом аспекте. Перспективные направления включают: использование полученных данных для разработки рекомендательных систем и мониторинга потребительских впечатлений (*consumer sentiment tracking*); проведение сравнительного междисциплинарного анализа визуальных и лингвистических маркеров эмоций для выявления универсальных и культурно-специфических паттернов экспрессии; включение экспериментальных данных о восприятии контента аудиторией для проверки гипотез об эмоциональном воздействии рассмотренных эмоциональных арок и повышения валидности интерпретаций.

ЛИТЕРАТУРА

1. Liu B. Sentiment analysis and subjectivity // Handbook of natural language processing / ed. by Nitin Indurkha, Fred J. Damerau. 2nd ed. Boca Raton: Chapman and Hall/CRC, 2010. P. 40.

2. Smetanin S. The applications of sentiment analysis for Russian language texts: Current challenges and future perspectives // IEEE Access. 2020. Vol. 8. P. 110693–110719.
3. Darwin C. The expression of emotion in animals and man. London, 1877.
4. Ekman P. An argument for basic emotions // Cognition and Emotion. 1992. Vol. 6, No. 3–4. P. 169–200. DOI: <https://doi.org/10.1080/02699939208411068>.
5. Plutchik R. A general psychoevolutionary theory of emotion // In Plutchik R., Kellerman H. (Eds.), Emotion: Theory, research, and experience: Vol. 1. Theories of emotion. Academic Press, 1980. P. 3–33.
6. Russell J. A. A circumplex model of affect // Journal of Personality and Social Psychology. 1980. Vol. 39, No. 6. P. 1161–1178. DOI: <https://doi.org/10.1037/h0077714>.
7. Kim E., Klinger R. A survey on sentiment and emotion analysis for computational literary studies // arXiv preprint arXiv:1808.03137. 2018. URL: <https://arxiv.org/abs/1808.03137>.
8. Mayer J. D., Roberts R. D., Barsade S. G. Human abilities: Emotional intelligence // Annual Review of Psychology. 2008. Vol. 59. P. 507–536. DOI: <https://doi.org/10.1146/annurev.psych.59.103006.093646>.
9. Ziora L. The sentiment analysis as a tool of business analytics in contemporary organizations // Studia Ekonomiczne. 2016. Vol. 281. P. 234–241.
10. Rambocas M., Pacheco B. G. Online sentiment analysis in marketing research: a review // Journal of Research in Interactive Marketing. 2018. Vol. 12, No. 2. P. 146–163.
11. Bougie R., Pieters R., Zeelenberg M. Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services // Journal of the Academy of Marketing Science. 2003. Vol. 31, No. 4. P. 377–393.
12. Bollen J., Mao H., Zeng X. Twitter mood predicts the stock market // Journal of computational science. 2011. Vol. 2, No. 1. P. 1–8.
13. SentiHealth-Cancer: a sentiment analysis tool to help detecting mood of patients in online social networks / Rodrigues R. G., das Dores, R. M., Camilo-Junior, C. G., Rosa, T. C // International journal of medical informatics. 2016. Vol. 85, No. 1. P. 80–95.
14. Sharma U., Datta R. K., Pabreja K. Sentiment analysis and prediction of election results 2018 // Social Networking and Computational Intelligence. 2020. P. 727–739.
15. Samothrakis S., Fasli M. Emotional sentence annotation helps predict fiction genre // PLOS ONE. 2015. Vol. 10, No. 11. P. e0141922.
16. Heuser R., Moretti F., Steiner E. The emotions of London. Stanford, 2016.
17. Taboada M., Gillies M. A., McFetridge P. Sentiment classification techniques for tracking literary reputation // LREC workshop: Towards computational models of literary analysis. 2006. P. 36–43.

18. Nalisnick E. T., Baird H. S. Character-to-character sentiment analysis in shakespeare's plays // *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. 2013. Vol. 2. P. 479–483.
19. The emotional arcs of stories are dominated by six basic shapes / Reagan A. J., Mitchell L., Kiley D., Danforth C. M., Dodds P. S // *EPJ Data Science*. 2016. Vol. 5, No. 1. P. 1–12.
20. Multimodal Transformer for Unaligned Multimodal Language Sequences / Tsai, Y. H. H., Bai, S., Liang, P. P., Kolter, J. Z., Morency, L. P., Salakhutdinov, R // *Proceedings of the ACL*. 2019.
21. Wankhade M., Rao A. C. S., Kulkarni C. A survey on sentiment analysis methods, applications, and challenges // *Artificial Intelligence Review*. 2022. Vol. 55, № 7. P. 5731–5780.
22. Sentiment strength detection in short informal text / Thelwall M., Buckley K., Paltoglou G., Cai D., Kappas A // *Journal of the American Society for Information Science and Technology*. 2010. Vol. 61, № 12. P. 2544–2558.
23. Двойникова А. А., Карпов А. А. Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных // *Информационно-управляющие системы*. 2020. № 4 (107). С. 20–30.
24. Кирина М. А. Think about what you've learned: анализ тональности для моделирования пользовательского опыта в сфере онлайн-образования // *Человек: Образ и сущность. Гуманитарные аспекты*. 2024. № 2 (58). С. 176–204.
25. Devika M. D., Sunitha C., Ganesh A. Sentiment analysis: a comparative study on different approaches // *Procedia Computer Science*. 2016. Vol. 87. P. 44–49.
26. Das R., Singh T. D. Multimodal sentiment analysis: a survey of methods, trends, and challenges // *ACM Computing Surveys*. 2023. Vol. 55, № 13s. P. 1–38.
27. Kleinbaum D. G., Dietz K., Gail M., Klein M. *Logistic Regression*. Atlanta, GA, USA: Springer, 2002.
28. Gunn S. R. Support vector machines for classification and regression // *ISIS Tech. Rep.* 1998. Vol. 14, № 1. P. 5–16.
29. BERT: Pretraining of deep bidirectional transformers for language understanding / Devlin J., Chang M. W., Lee K., Toutanova K // *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol. (NAACL HLT)*. 2019. Vol. 1. P. 4171–4186.
30. Distributed representations of words and phrases and their compositionality / Mikolov T., Sutskever I., Chen K., Corrado G., Dean J // *Proc. 26th Int. Conf. Neural Inf. Process. Syst. (NIPS)*. 2013. Vol. 2. P. 3111–3119.
31. Pennington J., Socher R., Manning C. D. Glove: Global vectors for word representation // *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*. 2014. P. 1532–1543.
32. Duppada V., Jain R., Hiray S. SeerNet at SemEval-2018 task 1: Domain adaptation for affect in tweets // *Proc. 12th Int. Workshop Semantic Eval*. 2018. P. 18–23.
33. Konig A. C., Brill E. Reducing the human overhead in text categorization // *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 2006. P. 598–603.

34. Lu K., Wu J. Sentiment analysis of film review texts based on sentiment dictionary and SVM // *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence*. 2019. P. 73–77.
35. Baid P., Gupta A., Chaplot N. Sentiment analysis of movie reviews using machine learning techniques // *International Journal of Computer Applications*. 2017. Vol. 179, No. 7. P. 45–49.
36. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis / Rehman A. U., Malik A. K., Raza B., Ali W // *Multimedia Tools and Applications*. 2019. Vol. 78. P. 26597–26613
37. Bhoir P., Kolte S. Sentiment analysis of movie reviews using lexicon approach // *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. 2015. P. 1–6.
38. Chu E., Roy D. Audio-visual sentiment analysis for learning emotional arcs in movies // *2017 IEEE International Conference on Data Mining (ICDM)*. 2017. P. 829–834.
39. Plantinga C., Smith G. M. *Passionate views: Film, cognition, and emotion*. Baltimore: Johns Hopkins University Press, 1999.
40. Hipson W. E., Mohammad S. M. Emotion dynamics in movie dialogues // *PLOS ONE*. 2021. Vol. 16, No. 9. P. 1–19.
41. Improving productivity in Hollywood with data science: Using emotional arcs of movies to drive product and service innovation in entertainment industries / Del Vecchio M., Kharlamov A., Parry G., Pogrebna G // *Journal of the Operational Research Society*. 2021. Vol. 72, No. 5. P. 1110–1137.
42. Salinas-Muñoz C., Stange-Marcus H. Experimentando las modernidades latinoamericanas: Estrategias melodramáticas en los filmes *La vendedora de rosas* y *Pizza, birra, faso* // *Revista Austral de Ciencias Sociales*. 2015. № 29. P. 165–179.
43. Cappello G. *Una ficción desbordada. Narrativa y teleseries*. Lima: Fondo Editorial de la Universidad de Lima, 2015.
44. Gledhill C. *The Melodramatic Field: An Investigation* // *Film Criticism*. 1985. Vol. 9, № 1. P. 5–27.
45. Hanich J., Menninghaus W., Wilder S. Beyond sadness: The multi-emotional trajectory of melodrama // *Cinema Journal*. 2017. Vol. 56, № 4. P. 76–101.
46. Колмогорова А. В., Калинин А. А., Маликова А. В. Лингвистические принципы и методы компьютерной лингвистики для решения задач сентимент-анализа русскоязычных текстов // *Актуальные проблемы филологии и педагогической лингвистики*. 2018. № 1 (29). С. 139–148
47. Hartmann J. *Emotion English DistilRoBERTa-base* [Electronic resource] // Hugging Face. 2022. URL: <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/>
48. Liu H., Lieberman H., Selker T. A model of textual affect sensing using real-world knowledge // *Proc. ACM Conf. Intell. User Interfaces*. 2003. P. 125–132.

49. Alm C. O., Roth D., Sproat R. Emotions from text: machine learning for text-based emotion prediction // Proc. Hum. Lang. Technol. Conf. Conf. Empir. Methods Natural Lang. Process. (HLT/EMNLP). 2005. P. 579–586.
50. Diale T. Sentiment analysis of social media data using Supervised learning and Ekman's basic emotions theory: PhD дис. Gaborone, Botswana, 2018.
51. The emotional arcs of stories are dominated by six basic shapes / Reagan A. J., Mitchell L., Kiley D., Danforth C. M., Dodds P. S. // EPJ Data Science. 2016. Vol. 5, No. 1. P. 31. DOI: <https://doi.org/10.1140/epjds/s13688-016-0093-1>
52. Дудина Т. П. Драматические жанры. Федеральное государственное бюджетное образовательное учреждение высшего образования «Елецкий государственный университет им. И. А. Бунина», 2019.
53. Frome J. Melodrama and the Psychology of Tears // Projections. 2014. Vol. 8, No.1. P. 23–40.

УДК 78:004.912 + 004.9::008 + 81'366.3 + 81'25

Кочкурова Анастасия Егоровна

студент департамента филологии
Научный исследовательский
университет «Высшая школа
экономики»
г. Санкт-Петербург, Россия

Anastasia Kochkurova

Student of the Department of Philology
National Research University «Higher
School of Economics»
St. Petersburg, Russia
anastasia5855347nk@gmail.com

СЕНТИМЕНТ-АНАЛИЗ КАК ИНСТРУМЕНТ ИССЛЕДОВАНИЯ МУЗЫКАЛЬНОГО ТЕКСТА (на материале песен группы «Мельница»)

В рамках данного исследования с помощью метода компьютерной лингвистики, сентимент-анализа, позволяющего определять эмоциональную окраску текста, проводится попытка определения лексической тональности песен русскоязычной фольк-рок группы «Мельница». Описываются трудности, возникающие при обработке русскоязычных текстов с высокой степенью художественности, образности. В качестве возможного решения предлагается проанализировать композиции с помощью кастомизированных словарей, составленных на основе частотного словаря и экспериментального опроса. Выдвигается гипотеза, что творчество группы «Мельница» обладает сильной положительной лексической эмоциональностью. Для проверки предположения используется сочетание компьютерных технологий и методов перцептивного анализа. В заключение делается вывод о возможностях использования метода расширения словарей в сентимент-анализе применительно к музыкальным текстам.

Ключевые слова: обработка музыкального текста; сентимент-анализ; анализ перцепции; бинарная классификация; digital humanities.

SENTIMENT ANALYSIS AS A TOOL FOR STUDYING MUSICAL TEXTS: A CASE STUDY OF THE BAND "MELNITSA"

This study employs the method of computational linguistics, sentiment analysis, which determines the emotional coloring of text, in an attempt to analyze the lexical tone within the songs of the Russian folk-rock band "Melnitsa". The difficulties encountered when processing Russian-language texts characterized by a high degree of artistry and figurativeness (poetic language) are described. As a potential solution, the analysis of the compositions is proposed using customized dictionaries compiled based on a frequency dictionary and an experimental survey. The hypothesis is put forward that the creative output of the band "Melnitsa" exhibits strong positive lexical emotionality. To test this assumption, a combination of computer technologies and perceptual analysis methods is utilized. In conclusion, the study draws inferences regarding the potential of using dictionary expansion methods in sentiment analysis as applied to musical texts.

Key words: musical text processing; sentiment analysis; perceptual analysis; binary classification; digital humanities.

Анализ тональности или сентимент-анализ – класс методов, позволяющих определить эмоциональную составляющую текста с помощью инструментов автоматической обработки естественного языка. Поле исследований, в рамках которых используется сентимент-анализ, преимущественно сосредоточено в отрасли маркетинга, где метод активно применяется для выявления отношения покупателя к товару [1]. Однако определение тональности не ограничивается маркетингом и, как показывают современные работы [2], может использоваться в других целях, в том числе для анализа литературных произведений или музыкальных композиций [3]. В области литературоведения анализ тональности может помочь для определения идиостиля писателя [4], тогда как музыковеды активно используют технологии распознавания эмоций для создания рекомендательных систем [5], изучения пользовательского опыта [6] или же для характеристики музыкальных тенденций [7].

Преимущество технологии сентимент-анализа состоит в том, что за короткий промежуток времени такой подход позволяет проанализировать большой объем данных, который в последствии можно использовать как для характеристики творчества группы, так и для анализа восприятия слушателя. Недостаток метода состоит в ограниченной возможности его использования: фокус на обработке текстов социальных медиа, отзывов, комментариев и прочего пользовательского контента приводит к тому, что модели сентимент-анализа преимущественно обучаются на датасетах или словарях, собранных из текстов соответствующего типа. По этой причине, когда исследователь применяет готовый инструментарий для оценки тональности, он может столкнуться с искаженными значениями, которые могут противоречить общему смыслу произведения. Особенную актуальность эта проблема приобретает при анализе текстов с высокой степенью образности, где лексическое наполнение выходит за рамки общеупотребительных единиц.

Одна из целей в настоящей работе – предложить стратегии для решения проблемы, направленные на расширение словарей с помощью перцептивного эксперимента и словаря частотной лексики на основе целевого корпуса. В качестве объекта исследования выбраны тексты фолк-роковой группы «Мельница». О высокой образности и мифологизации в текстах этой группы могут свидетельствовать исследования Е. К. Дробышевой [8], К. И. Одинцовой и М. В. Олюниной [9], посвященные анализу мифологем в творчестве группы. Так, например, исследователь А. А. Рыжков смог подтвердить гипотезу о том, что структурно (благодаря реконструкции пространства диптиха) тексты двух альбомов группы соответствуют характерной для мифологии модели «Мирового дерева» [10].

Кажется, образность лексем обуславливается самим фольклорным жанром, который может в определенной степени «диктовать» выбор языковых единиц. Так, в композиции «Ушба» помимо частотных «лед» и «небо», встре-

чаются также и «жатва», «лезвия», «пряха». Стоит отметить, что специфика текстов с высоким уровнем художественности может также подразумевать трансформацию тональности слова – процесс, при котором частотная единица, например, то же существительное «лед», приобретает эмоциональную коннотацию.

В интервью радиостанции «Эхо Перми» солистка группы, Наталья Андреевна О'Шей, отмечала, что творчество определяется целебной функцией: «Творчество должно помогать сублимировать какие-то проблемы, какие-то вещи, которые людей беспокоят, пусть это будет даже какой-то болезненный катарсис, но мне кажется, что искусство должно быть именно таким» [11]. Данный взгляд на цели искусства, вероятно, лежит в основе творческого подхода «Мельницы» и может проявляться в стремлении к положительному разрешению внутренних конфликтов через прослушивание композиций. Ключевая гипотеза данного исследования состоит в том, что для группы «Мельница» характерна положительная эмоциональность текстов. Таким образом, **второй целью** настоящей работы является анализ эмоциональности в текстах песен группы «Мельница». Для реализации указанных целей потребовалось несколько этапов.

Этапы обработки текста

На первом этапе работы была просмотрена дискография «Мельницы», выделены периоды, в которые выходили альбомы. С помощью парсера «Web Scraper» был составлен корпус текстов песен, в который вошли 11 альбомов во временном промежутке с 2002 по 2023 год. Суммарный объем корпуса составил 11 675 токенов. Для работы с корпусом было выбрано приложение «Orange» [12], предоставляющее необходимые инструменты для sentiment-анализа русскоязычных текстов. На первом этапе был загружен 121 файл, каждый из которых представлял одну композицию в формате txt. После чего файлы были предобработаны (приведение к нижнему регистру, токенизация, лемматизация с помощью «Lemmagen Lemmatizer»). С помощью встроенных словарей оценочной лексики «Multilingual sentiment lexicons» от Data Science Lab был проведен sentiment-анализ. Использованный пайплайн представлен на Рисунке 1.



Рис. 1. Путь обработки данных в приложении

Промежуточные результаты

Инструментарий приложения «Orange» позволяет проанализировать тональность по правилам суммирования, при которых каждое слово из документа сопоставляется с множеством из положительного и отрицательного словарей, после чего рассчитывается разница между положительными и отрицательными словами, значение нормализуется путем его деления на длину текста и умножением на 100.

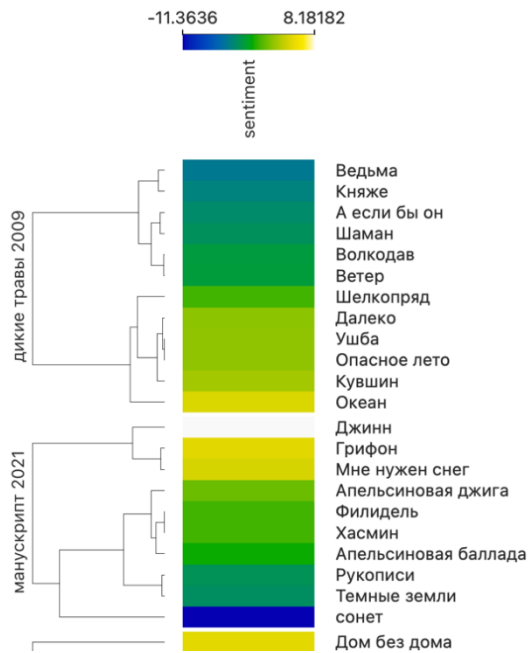


Рис. 2. Первоначальные значения сентимента на альбом

На первом этапе работы со словарями положительной и отрицательной лексики стало очевидно, что результаты выдачи не отражают лексических особенностей композиций группы (Рисунок 2). Так, например, для песни «Сонет», отрицательная тональность составила – 11.3636, максимальное значение по шкале тональности. Результат такой разметки выделялся на фоне других композиций, например, песни «Темные земли», которая, несмотря на устойчивый образ смерти, парадоксально оказалась более положительной, чем сонет, в котором прослеживалась не только печаль, но и нежность.

Причиной искажения результатов стал взятый за основу анализа словарь: он содержал частицы, слова на английском языке, существительные без явной эмоциональной окраски. Так, например, в словарях были замечены следующие лексемы: революционер(pos), как(pos), так(pos), well(pos), во (neg), комплекс(neg), информация(neg). После проверки с помощью одного из самых крупных словарей эмоциональной лексики («RuSentiLex»), было установлено, что классификация тональности по предлогам возможна только в контексте, в противном случае, она нерелевантна. Таким образом, было принято решение отредактировать использованные словари.

Расширение словаря за счет частотного списка

Для решения задач по возможной кастомизации, персонализации словаря был применен метод, описанный в методических рекомендациях по работе с приложением SentiStrength [13], где авторы обосновывали добавление слова к позитивной или отрицательной группе на основе частотного словаря. По аналогии был составлен частотный словарь, с помощью виджета «word cloud», а также расширен список стоп-слов.



Рис. 3. Облако слов до обновления стоп-словаря

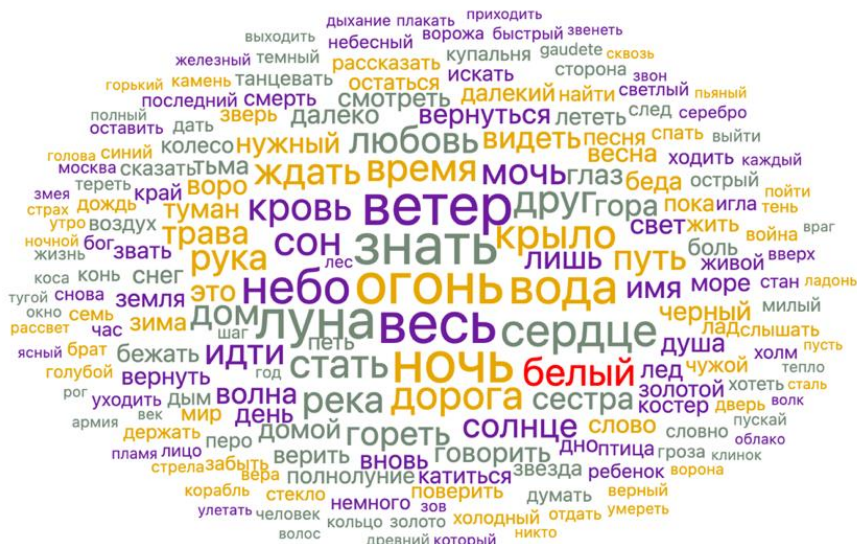


Рис. 4. Облако слов после обновления стоп-словаря

В итоговую выборку вошли следующие лексемы: «Огонь», «Сердце», «Солнце», «Белый», «Черный», «Лед», «Война», «Нужный», «Последний». Слова были выбраны как по частоте, так и по возможному наличию тональности. Однако помещать лексемы в словари на данном этапе, ссылаясь при этом на собственный опыт, было бы опрометчиво и субъективно, поэтому было принято решение провести опрос по разметке лексем в текстах, что позволило бы снять субъективность и не ограничиваться мнением одного человека.

Кастомизация словаря с помощью опроса

Расширение словаря осуществлялось на основе проведения опроса с использованием специально разработанного для этой задачи чат-бота на платформе «Telegram». Использование такого способа опроса было обусловлено следующими причинами.

1. **Быстрота прохождения.** Респондентам не приходилось переходить на следующую вкладку при выборе ответа. Бот сам выводил следующий вопрос. Задача респондентов сводилась к тому, чтобы прочитать вопрос и нажать на правильную кнопку. При таких условиях время прохождения занимало около пяти минут.

2. **Непрерывный порядок.** Вопросы подавались респондентам один раз. Они не могли вернуться к предыдущим ответам, не могли и соотнести вариант ответа с последующим, так как не видели список вопросов.

Бот состоял из приветствия, инструкций и четырех траекторий, помеченных цветом: зеленым, синим, фиолетовым и желтым. На каждый сет приходилось 20 вопросов. Длина кода составила 949 строк.

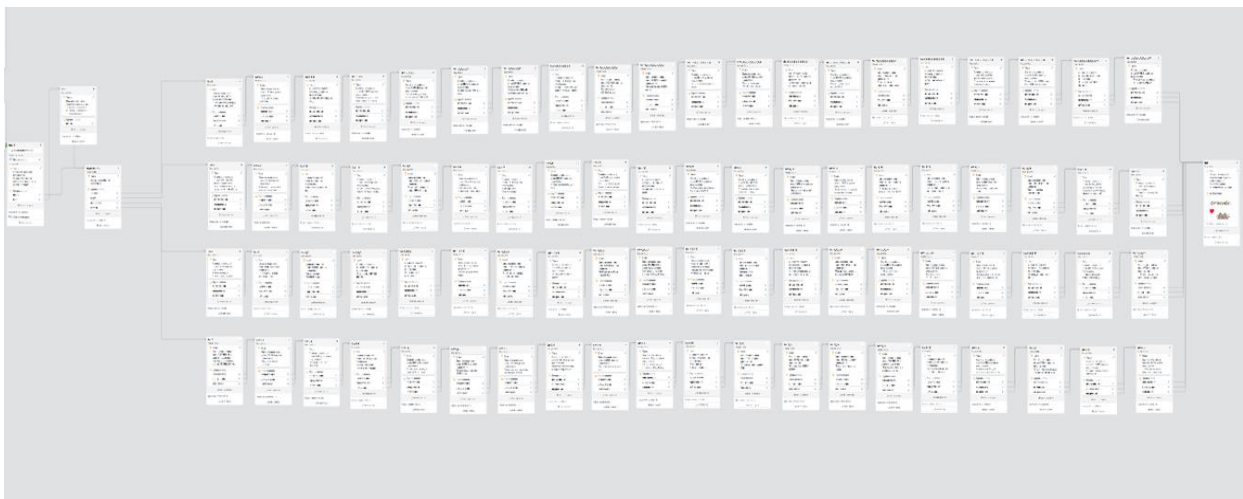


Рис. 5. Структура всего бота, выведенная с помощью виджета «G-grapf»

На каждой карточке было указано прописными буквами нужное слово, к нему с помощью «Orange» был подобран соответствующий контекст (Рисунок 6), для того чтобы респонденты не анализировали слово в отрыве от текста. Пример карточки можно увидеть на рисунке 7. Изначально выбирались такие примеры с контекстом, чтобы содержание считывалось насколько это возможно, учитывая образность фолк-текстов. Под контекстом здесь понимается предложение от точки до точки, в которое входит слово, также к нему могла добавиться либо строка «до», либо строка «после».

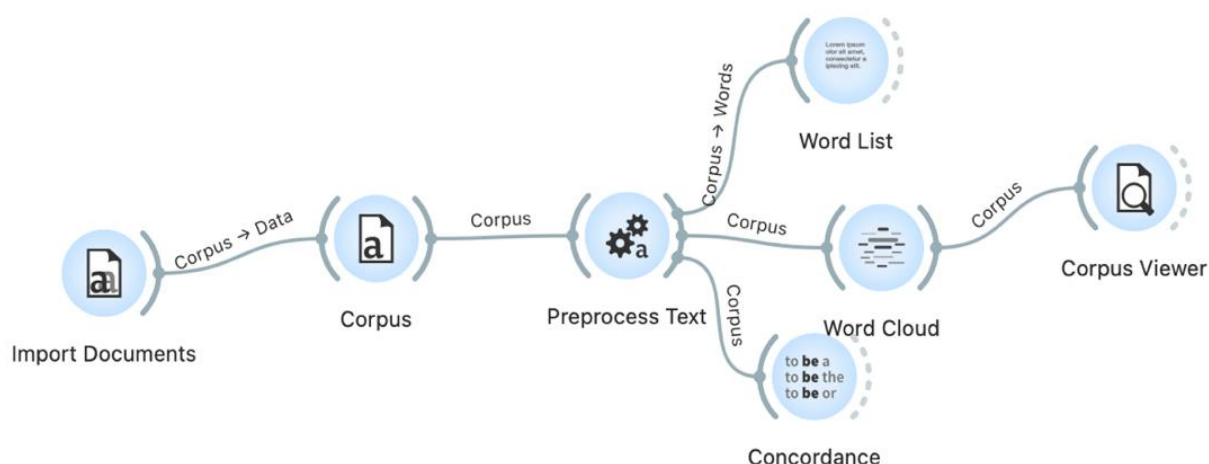


Рис. 6. Поиск контекстов в «Orange»

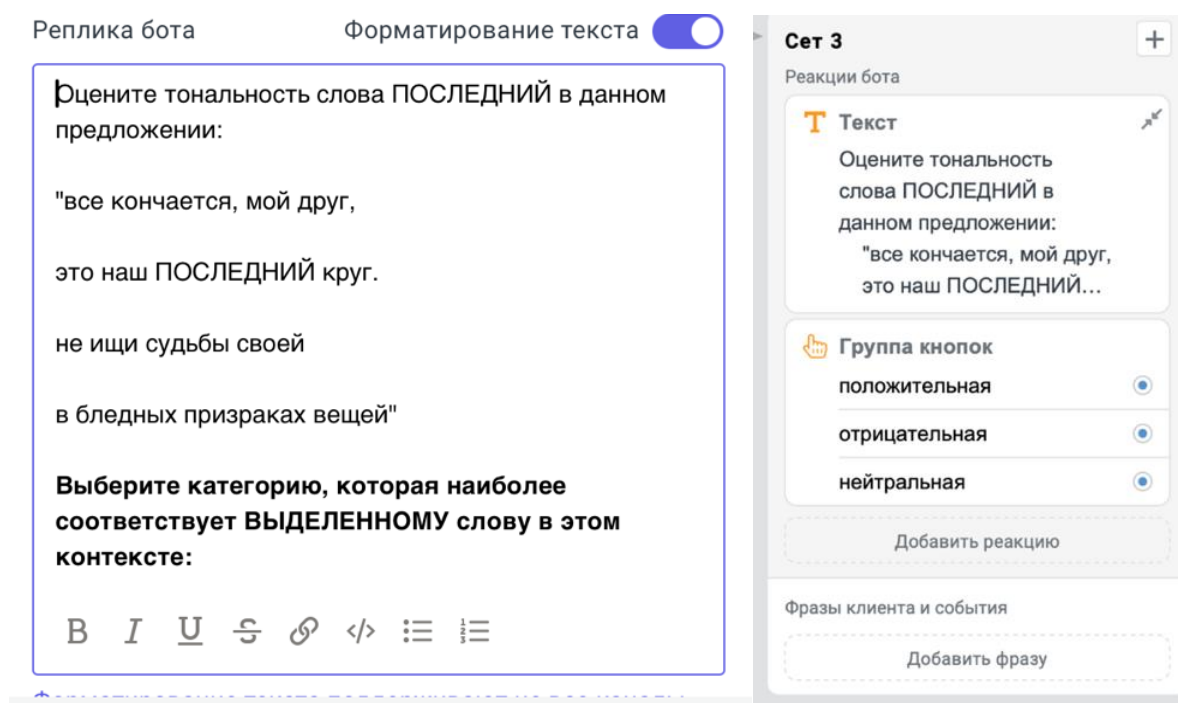


Рис. 7. Внешний вид карточки через виджет «G-graph»

Для каждого вопроса респондентам было предложено, отталкиваясь от своих эмоций, оценить тональность слова в контексте как «положительную», «отрицательную» или «нейтральную». В инструкции было прописано, что подразумевалось под каждым из видов тональностей. *Положительная тональность* означала, что слово вызвало приятные эмоции такие как, например: радость, надежда, восхищение, удовлетворение, облегчение. *Отрицательная тональность* – слово оставило неприятный осадок, вызвало негативные эмоции, такие как, например, горе, страх, негодование, отвращение, жалость, разочарование. *Нейтральная тональность* – слово не вызвало перечисленных ранее эмоций.

Разделение на три тональности вызвано необходимостью добавлять размеченные слова только в два словаря с положительной и отрицательной лексикой. Слова, которые респонденты отмечали в качестве «нейтральных», считались недостаточно эмоциональными и в словарь не включались.

Результаты разметки

В опросе приняли участие 141 респондент. Для наглядной репрезентации данных была составлена таблица с подсчитанными процентами «положительной», «отрицательной» и «нейтральной» тональности для слова.

Таблица 1

Результаты разметки

Слово	Положительная		Отрицательная		Нейтральная		Всего
	Кол-во	%	Кол-во	%	Кол-во	%	
Огонь	258	68,25	60	15,87	60	15,87	378
Сердце	122	47,10	40	15,44	97	37,45	259
Белый	160	37,21	120	27,91	150	34,88	430
Солнце	166	59,71	55	19,78	57	20,50	278
Нужный	131	65,83	26	13,07	42	21,11	199
Черный	43	14,73	116	39,73	133	45,55	292
Лед	24	12,70	109	57,67	56	29,63	189
Война	24	9,13	192	73,00	47	17,87	263
Последний	76	25,25	139	46,18	86	28,57	301

По результатам разметки были сделаны следующие выводы для частотных лексем.

1. **«Огонь».** Как видно по частотному словарю (см. Рисунок 4) огонь в творчестве группы «Мельница» является самым популярным существительным. Любопытно также, что именно «огонь» оказался самым положительным словом по разметке. Возможно, респонденты размечали огонь как нечто положительное, потому что считывали в нем своеобразную функцию «помощника»:

*Немного ОГНЯ – середина пути,
Немного ОГНЯ тебя может спасти
В блеске обмана, обмана*
(Мельница. Немного огня)

Предположительно, в творчестве «Мельницы» огонь выполняет функцию, которая определяется в мифе как «проживание героем процесса инициации»: «Я иду – через лето иду к тебе по горячим шпалам. / Каждый шаг – дорога в ОГОНЬ, дорога к ОГНЮ» (Мельница. Дорога в огонь). Однако для проверки гипотез требуются дальнейшие исследования.

2. **«Лед».** Данная лексема рассматривалась в рамках антитезы огонь-лед. В отличие от огня, лед оказался одним из самых отрицательных слов в разметке. Стоит отметить, что контекстуальное окружение льда включает в себя такие лексемы как: «больно», «жжет», «кровоточат», что предположительно могло сформировать образ негативного явления, которое ранит лирического героя/героиню.

3. **«Нужный».** Изначальная гипотеза предполагала, что данная лексема не обладает тональностью. С одной стороны, потребность что-то иметь может отражаться в позитивном стремлении, с другой стороны она же может свидетельствовать и о недостатке. Словарь встроенной негативной лексики Multilingual sentiment lexicons, по-видимому, придерживался второй точки зрения: в нем нужда была включена в негативный список. Результаты исследования, однако, выявили преобладание позитивной тональности (более 60 % оценок респондентов).

4. **«Война».** Отсутствие существительного война в негативном словаре «Orange» при наличии производных («военный», «танк») представлялось проблемой. Включение могло бы быть обоснованно деривационной связью, однако лексема была избрана в качестве «контрольного элемента» с заведомо ожидаемой негативной коннотацией. Такой выбор позволял оценить осознанность респондентов при прохождении эксперимента. Результаты подтвердили ожидания: лексема получила однозначно негативную маркировку, что свидетельствует о валидности ответов.

5. **«Последний».** Несмотря на 20 % респондентов, считающих тональность лексемы положительной, типичное окружение предсказывает отрицательный контекст «Уже стремится в тот ПОСЛЕДНИЙ рейс, / Где мы с тобой встретились...» (Мельница. Марсианский экспресс); «Все кончается, мой друг, / это наш ПОСЛЕДНИЙ круг.» (Мельница. Мертвец); «ПОСЛЕДНИЙ весенний блюз. Экзюпери лениво гадает – дотяну или разобьюсь?» (Мельница st. Exupéry Blues).

Несмотря на доминирующую семантику конца и утраты, полученные двадцать процентов позволяют предположить наличие специфического терапевтического эффекта или восприятия облегчения, которое связано с завершением.

6. **«Белый / Черный».** На этапе выбора данных лексем была поставлена цель проверить тональность у цветовых характеристик. Результаты показали преимущественно нейтральное восприятие. Гипотетически, это

может объясняться высокой степенью образности фолк-текстов, где живые образы и метафоры доминируют над статичным употреблением цвета, снижая его эмоциональную нагрузку.

7. «Солнце». Выбранное существительное также являлось контрольным элементом, позволяющим судить о репрезентативности данных. Однако стоит отметить, что в опрос были добавлены не только явно положительные примеры, такие как: «К востоку рвутся вороны / Средь облачных зыбей, / Но СОЛНЦЕ – князь с коронною / Терновою сверхновою – / Спешит в другую сторону, / К тебе, к тебе, к тебе!» (Мельница. Анестезия), но и неоднозначные, например: «Полыхнуло СОЛНЦЕ / Зов крови / Давно ли ты понял / что никто не вернется» (Мельница. Зов крови). Несмотря на это солнце все же получила статистически значимую положительную тональность, и было включено в словарь.

8. «Сердце». Существительное было отобрано по причине его высокой частотности. Для оценки были подобраны разные примеры, с предположительно отрицательной или положительной тональностями. По результатам опроса большинство респондентов разметило лексему как положительную.

По наименьшему значению в столбце «Всего» заметно, что минимальный порог разметки каждого слова составил 189. Таким образом, чтобы слово вошло в словарь, ему нужно было преодолеть установленный порог в 46 процентов. Как итог в положительном словаре оказались: **огонь**, **сердце**, **солнце**, **нужный**. В отрицательный словарь: **лед**, **война**, **последний**. Не вошли в словарь цвета: белый и черный.

После добавления указанных слов работа со словарями продолжилась: из списков была удалена английская лексика, однако для сохранения связи с встроенными словарями удаленные слова заменялись русскими эквивалентами. Из словарей исключались также и лексемы, которые, возможно, подошли бы для анализа текстов социальных сетей, но при работе с художественной лексикой могли вызвать вопросы и могли привести к ошибочным результатам. Так, из словаря с негативной лексикой был удален «ребенок», а из положительного – «жертвенный». Далее в готовый словарь была добавлена предварительно почищенная выдача из русскоязычного «Русентилекса». После объединения двух словарей были оставлены только уникальные слова, т. е. удалены повторы. Затем полученные значения сентимента визуализировались с помощью виджета «Heat map» (см. рисунок 8).



Рис. 8. Конечный путь обработки данных в приложении Orange

Результаты кастомизации словаря

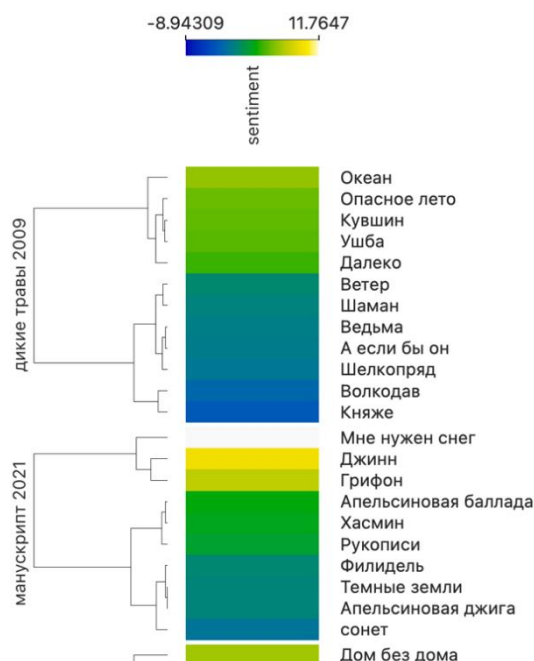


Рис. 9. Пример итоговых значений сентимента

Как показывает «Orange», кастомизация словаря оказалась успешной. Заметно, что значение сентимента сгладилось: песни, которые считывались изначально как явно негативные, оказались по середине шкалы. Так, например, изначально размеченная ярко-синим (см. рисунок 2) песня «Сонет» теперь приобрела лазурный оттенок, что значит, стала менее отрицательной по разметке. Такие результаты доказывают возможность разметки с помощью частотного словаря и эффективность проведения опроса.

Для подтверждения гипотезы о том, что творчество группы «Мельница» обладает положительной эмоциональностью, числовые показатели каждой песни были суммированы и объединены по принципу альбома. Использование метода суммирования

показалось релевантным в данном контексте, так как основная задача была не столько установить усредненное числовое значение, сколько выявить его положительный или отрицательный характер. Сумма же позволила отталкиваться от подавляющего числа положительных или отрицательных песен.

Таблица 2

Распределение сентимента на альбом

Название композиции	Сентимент
Символ солнца	1,17
Манускрипт	26, 64
Люцифераза	20,59
Алхимия	17,01
Химера	16,41
Радость моя	6,67
Ангелофрения	12,48
Дикие травы	5,43
Зов крови	25,9
Перевал	6,16
Дорога сна	7,09

По таблице 2 заметно, что большая половина (шесть альбомов из одиннадцати) обладают сильным положительным сентиментом. По суммарному значению самыми положительными альбомами оказались «Манускрипт», «Зов крови», «Люцифераза». По количеству вхождений альбом «Манускрипт» (6 песен из 10), «Ангелофрения» (8 песен из 11) и «Алхимия» (8 песен из 12) – оказались лидерами по положительной тональности. «Дикие травы» (7 песен из 11) и «Символ солнца» (5 песен из 12) – лидеры по количеству песен с отрицательной тональностью. Напомним, что под отрицательной тональностью в данном случае понимается значение ниже нуля, под положительной – выше нуля. Что касается результатов на уровне композиций, то здесь наибольший показатель отрицательной тональности имеет песня «Воин вереска» из альбома «Дорога сна». Если обратиться к тексту, то можно увидеть, что образный ряд и метафоры: топос холода и мглы; воин, который никогда не вернется и просит не лить по нем слезы, – подтверждают такую классификацию.

*И не сомкнуть кольцо седых холмов,
И узок путь по лезвию дождя,
И не ищи – ты не найдешь следов,
Что Воин Вереска оставил, уходя.*

(Мельница. Воин Вереска)

Самая положительная песня по анализу Orange – «Мне нужен снег». Здесь, как кажется, сыграло роль то, что «нужно» было включено в словарь положительных слов.

*Если зима – зима, мне нужен снег
(Мне срочно нужен снег),
Чтоб хотя бы слепить снеговика.*

(Мельница. Мне нужен снег)

На втором месте композиция тоже из альбома «Манускрипт» Джинн.

*Слушай же, мне нужен джинн,
О, как мне нужен джинн, о великий царь,
Самый проворный джинн, ведь мне так нужно домой,
Мне так нужно домой – к тебе.*

(Мельница. Джинн)

Таким образом, по результатам анализа наибольшее модульное значение на шкале (см. рисунок 10) получила положительная тональность, что означает, подавляющее количество слов в текстах «Мельницы» окрашены положительно. Также по итоговой таблице 2 заметно, что числовые показатели альбомов высокие и положительные, следовательно, лексический сентимент песен положительный. Указанные выводы подтверждают гипотезу о положительной эмоциональности, которая свойственна песням группы «Мельница».

Общие выводы

На сегодняшний день в области сентимент-анализа нет доступных функций, позволяющих с одинаковой точностью оценить тенденции как в маркетинге, так и в художественном тексте. Однако эффективным решением для преодоления ограниченного инструментария приложений сентимент-анализа может стать кастомизация, преобразование словарей с помощью частотных списков и опросов среди аудитории. Благодаря сочетанию двух подходов, одному из области digital humanities, другому из области психологии и перцептивных экспериментов, появляется возможность адаптировать словари к контекстам и добиться качественных результатов при анализе текстового материала.

ЛИТЕРАТУРА

1. Богданов А. Л., Дуля И. С. Сентимент-анализ коротких русскоязычных текстов в социальных медиа // Вестник Томского гос. ун-та. Экономика. 2019. № 47. С. 220–240.
2. Hu Q., Murad M., Li Q. Advancing music emotion recognition: large-scale dataset construction and evaluator impact analysis. Multimedia Systems, 2025.
3. Аристова А. И., Разоренова Ю. А. Основы теории эмоциональности художественного текста // Научное сообщество студентов XXI столетия. Гуманитарные науки. СибАК, 2017. С. 17–23.
4. Егорова А. Д., Станкевич Н. В. Сентиментные траектории как инструмент анализа текста // Культура и текст. 2024. №2 (57). С. 169–180.
5. Agres K. R., Schaefer R. S., Volk A., S. van Hooren, Holzapfel A., Dalla Bella S., Müller M., De Witte M., Herremans D., Ramirez Melendez R. et al., Music, computing, and health: a roadmap for the current and future roles of music technology for health care and well-being. Music & Science, vol. 4. 2021.
6. Dogaru I., Furnham A., McClelland A. Understanding how the presence of music in advertisements influences consumer behaviour. Acta Psychologica 248 (2024), 2024.
7. Napier K., Shamir L. Quantitative Sentiment Analysis of Lyrics in Popular Music // Journal of Popular Music Studies. 2018. P. 161–176.
8. Дробышева Е. К. Мифопоэтический компонент и его функции в лирике А. Васильева и Н. О`шей // Сборник научных статей VII Международной научно-методической конференции. М., 2019.
9. Одинцова К. И., Олюнина М. В. Образный ряд в песне «Хамсин» группы «Мельница» // Материалы VII всероссийской научно-практической конференции М., 2022. С. 22.
10. Рыжков А. А. Мифологическая модель пространства в текстах рок-группы «Мельница» (на материале альбомов «Алхимия» и «Химера»). вып. квалиф. работа. Челябинск. 2019.

11. «Современный фолк-рок» [Электронный ресурс] / радиостанция «Эхо Перми». 23.03.2021. Режим доступа: <https://echoperm.ru/interview/299/91196/> (дата обращения: 12.07.2025).
12. Orange Data Mining, версия 3.36.0 [Электронный ресурс]. Режим доступа: <https://orangedatamining.com/> (дата обращения: 12.07.2025).
13. Кольцов С. Н., Павлова Ю., Кольцова О. Ю. Метод автоматического анализа тональности текста в применении к социологическим задачам [Электронный ресурс]: методическое пособие. Режим доступа: <https://linis.hse.ru/sentiment-analysis-tekstov-postov-v-zhzh/> (дата обращения: 12.07.2025).

УДК 7.046:[004.738.5:82](510)

Ли Сюэсинь

соискатель культурологии,
аспирантка, соискатель кафедры
культурологии учреждения
образования БГУКИ
Белорусский государственный
университет культуры и искусств
г. Минск, Беларусь

Li Xuexin

Candidate of Cultural Studies,
Postgraduate Student, Applicant
of the Department of Cultural Studies
of the educational institution BSUKI
Belarusian State University of Culture
and Arts
Minsk, Belarus
lixuexin0828@gmail.com

ВОПЛОЩЕНИЕ МИФООБРАЗОВ В СЕТЕРАТУРЕ КИТАЯ

В статье исследуются механизмы формирования мифологических образов в китайской сетературе и их культурная специфика, уделяя внимание инновационной реконструкции на основе традиционных мифов. Древняя мифология представляет собой ценнейший пласт традиционной китайской культуры. В современной сетевой литературе, особенно в жанре сюаньхуань, происходит систематическая реконструкция фантастической вселенной, обладающей ярко выраженной китайской спецификой. Посредством тщательно разработанной мифопэтики данные произведения формируют разветвленную таксономию божественных и демонических сущностей, что способствует реинтеграции традиционных мифологических парадигм в пространство современного литературного дискурса. Авторы онлайн-литературы демонстрируют, как синтез китайских и западных мифологических традиций с современными ценностями создают новые нарративные модели, отвечающие запросам глобальной аудитории. Успех этих произведений на международном рынке показывает их место в культурном экспорте Китая, раскрывая потенциал сетевой литературы как медиатора межкультурного диалога. Подобная трансформация не только сохраняет символичность мифа, но и адаптирует его к эстетическим запросам цифровой эпохи, формируя уникальную парадигму кроссмедийного повествования.

Ключевые слова: миф; образ; сетература; Китай; парадигма.

THE EMBODIMENT OF MYTH IMAGES IN CHINESE ONLINE LITERATURE

The article examines the mechanisms of formation of mythological images in Chinese online literature and their cultural specificity, paying attention to innovative reconstruction based on traditional myths. Ancient mythology is a treasure of Chinese traditional culture. Modern online works present a fantasy world with Chinese characteristics, and gradually build a rich genealogy of gods and demons, making traditional mythology return to literature. Authors of online literature demonstrate how the synthesis of Chinese and Western mythological traditions with modern values creates new narrative models that meet the needs of a global audience. The success of these works in the international market emphasizes their role in China's cultural

export, revealing the potential of online literature as a mediator of intercivilizational dialogue. Such a transformation not only preserves the symbolism of the myth, but also adapts it to the aesthetic needs of the digital age, forming a unique paradigm of cross-media storytelling.

Key words: myth; image; online literature; China; paradigm.

Древние мифы являются сокровищем традиционной китайской культуры. В современную эпоху сильно фрагментированной информации использование сетературы является эффективным инструментом удовлетворения растущих духовных потребностей китайского народа. Развитие социальных сетей в Китае стало значимым социальным феноменом, который преобразил методы связи и обмена информацией под воздействием различных факторов. Это привело к созданию уникальной экосистемы социальных медиа, существенно изменившей формы общения между людьми. В свою очередь, социодинамика социальных сетей оказала глубокое влияние на различные сферы общества, включая бизнес и культуру. В результате взаимодействие в социальных сетях стало не только важным каналом коммуникации, но и мощным инструментом социального и экономического воздействия.

Социальные сети играют важнейшую роль в продвижении культурной продукции, предоставляя художникам, писателям, музыкантам и кинематографистам возможность делиться своим творчеством, делая его доступными для более широкой и разнообразной аудитории. Такое взаимодействие способствует формированию чувства общности и личной связи между субъектами творчества и их поклонниками, приводит к росту лояльности и поддержки. Пользователи социальных сетей часто делятся интересными или вдохновляющими их материалами.

В социальных сетях достойное место занимает сетература, которая начала динамично развиваться в конце XX в. По данным исследовательской группы «Доклад о развитии сетературы» Института литературы Китайской академии общественных наук, объем китайского рынка сетературы в 2024 г. достиг 5,07 млрд юаней, увеличившись на 16,5 % в годовом исчислении. Число писателей сетературы превысило 31,19 млн человек, число пользователей сетературы достигло 575 млн, а число зарубежных пользователей сетературы – 352 млн [1]. Сетевая культура с романами, комиксами, анимацией и играми с возвышенным мировоззрением в качестве основного носителя быстро проникает к пользователям Интернета через различные приложения для интернет-развлечений. Таким образом, в Китае развитие сетературы в наши дни получило взрывной старт. Литературные критики единодушно признают, что интернет-роман Цай Чжихэна «Первый интимный контакт» является основополагающим произведением китайской сетературы. Данное произведение имеет важное историческое значение в литературе, символизируя официальное зарождение китайской сетературы как новой ли-

тературной формы [2, с. 60]. Особенно быстро сетература начала распространяться с 2017 г., когда в Китае была создана крупнейшая платформа для чтения и литературного творчества Qidian (起点) – www.webnovel.com. Ее богатое содержание, технологии и капитал-резервы значительно детерминировали социодинамику китайской литературы в стране и за рубежом.

В последнее время в Китае наблюдается стремительный рост количества сетературы и дальнейшее увеличение численности ее авторов. В 2023 году в сфере создания сетературы наблюдалась заметная тенденция к омоложению, причем авторы, родившиеся после 1995 и 2000 годов, быстро выдвинулись на передний план, став основной силой в создании сетературы. Статистические данные платформы Tomato Novel (番茄小说), принадлежащей ByteDance (字节跳动), показывают, что количество авторов, присоединившихся к платформе в 2023 году, увеличилось на 42 % по сравнению с 2022 годом, при этом авторы, родившиеся после 1995 года, составляют 57 % от общего числа подписанных авторов, авторы, родившиеся после 1985 года, – 26 %, а авторы, родившиеся после 1975 года, – всего 9 % [1]. Так, фантастический роман «Гибель всего живого» (万族之劫), объемом 7,57 млн иероглифов молодого автора под псевдонимом Лаони Чи Сяоцзи (老鹰吃小鸡) стал самым востребованным у читателей. Книга лидировала по количеству подписок и комментариев: она собрала 3 млн постов в разделе обсуждений и принесла автору высокий доход. Собирают рекордное количество подписок и произведения других молодых писателей. Внушительных успехов добился автор с псевдонимом «Янь гуй чжэн чжуань» (言归正传). Его произведение «Мой старший просто слишком стабильный» (我的师兄实在是太稳健了), которое содержит в себе элементы магии, даосизма, мифологии и буддизма, возглавило список сетевых произведений по версии корпорации ChinaLiterature. Литературную премию имени Мао Дуня в 2019 г. получил Лян Сяошэнь (梁晓声) за роман «Мир человека» (人世间), объемом в 1,15 млн слов.

Современные онлайн-произведения представляют фэнтезийный мир с китайскими особенностями и постепенно выстраивают обширную генеалогию богов и демонов, тем самым возвращая традиционные мифы в сетературу. Несмотря на то, что современная научная рациональность заменила иррациональные верования в мифы, мифы в онлайн-литературных произведениях все еще пробуждают чувства и переживания людей, ведущие в мифический мир. Использование традиционной мифологии в онлайн-литературе многомерно. Наиболее распространенным является прямое присвоение мифологических терминов. Образы и черты характера мифологических персонажей, временные и пространственные регионы, инструменты и орудия труда в большом количестве интернет-романов напрямую взяты из таких произведений, как «Классика гор и морей», в которых записана древняя

мифология [3]. Прямое присвоение мифологических существительных максимизирует мифологический эффект. В отличие от прямого именования онлайн-фантастических романов, онлайн-исторические романы, как правило, создают мифологическую атмосферу, а обстановка персонажей больше похожа на многочисленных героев со странной внешностью в аутентичных мифах.

Авторы онлайн-литературы используют мифологию в качестве фона для украшения и построения текста. Взгляд на время и пространство, на жизнь и смерть, установленные традиционной китайской мифологией, потенциально предоставляют богатый материал авторам онлайн-романов [4, с. 14]. Возрождение традиционной мифологии в онлайн-литературе имеет свою собственную внутреннюю логику. Виртуальность самих онлайн-медиа определяет популярность вымышленных текстов, таких как фэнтези, сказки, боевые искусства и путешествия во времени. Среди четырнадцати основных категорий произведений на домашней странице китайского веб-сайта Qidian фэнтези, сказки и другие вымышленные типы составляют более половины. Этот вид вымышленной литературы имеет естественный спрос на воображение, и китайские и западные мифы естественным образом становятся повествовательными ресурсами для онлайн-литературы. У многих авторов онлайн-литературы собственного социального опыта и жизненного опыта недостаточно для поддержки литературного творчества. Поэтому они используют мифологические ресурсы для создания виртуального мира, чтобы восполнить потребности и трудности своего творчества. Причудливое воображение мифологии восполняет недостаток воображения писателей. Для читателей ожидание от чтения онлайн-статей заключается в получении эстетического удовольствия от них и снятии тревоги и напряжения в реальном мире. Прекрасного эмоционального опыта, содержащегося в мифологии, достаточно, чтобы восполнить различные недостатки и пустоту жизни в реальном мире. Как отмечал З. Фрейд: удовлетворение, вызванное реализацией подавленных желаний, настолько сильно, что оно определенно может компенсировать остаточные болезненные эмоции в течение дня [5]. В связи с этим сетература стала центром переписывания и воссоздания мифов, постепенно высвобождая больше возможностей между воспроизведением и деконструкцией.

Значение мифов в современных онлайн-фэнтези-романах заключается в выражении современных мыслей, основанных на традиционной культурной психологии, символическом или метафорическом способе. В пересказе национальных мифов онлайн-фэнтези-романы способны реконструировать национальную культурную память, усилить национальную идентичность и внутреннюю силу, и напряжение национального духа. Воссоздавая реалистичные мифы, авторы онлайн-фэнтези-романов подчеркивают разнообразную современность и индивидуальность. Современные онлайн-фэнтези-романы

основаны на культурной почве традиционных мифов и полагаются на появляющуюся коммуникационную среду современных сетей. В созданных сюрреалистических мифах они передают уверенность в национальной культуре и ценностях реальной жизни.

В последние годы фэнтезийные романы выделились из многих онлайн-литератур и стали категорией с наибольшим количеством авторов, наибольшим количеством текстов и наибольшей аудиторией. Эта разновидность романа стала новым типом романа, который объединяет метафизику, науку и литературу. С момента его возникновения расширение фэнтезийного романа постоянно увеличивалось. Все больше и больше авторов используют сюрреализм «Сюань» и магическую изменчивость «Сюань», чтобы создать красочный воображаемый мир для читателей [6, с. 70]. Мифы предоставляют их авторам сюжетные материалы и пространство для размышлений, что не только увеличивает культурную глубину онлайн-фэнтези-романов, но и расширяет реальность самих древних мифов.

В первой четверти нынешнего столетия все больше и больше онлайн-литературных произведений, содержащих китайскую мифологию Китая, начали выходить за границы. Онлайн-романы показывают фантастический мир с китайскими особенностями и постепенно выстраивают обширную и великолепную генеалогию мифических образов. В 2017 году Qidian Chinese запустил зарубежную версию Qidian International, что позволило расширить зарубежную читательскую аудиторию. Популярностью у зарубежных реципиентов пользуются такие онлайн-романы, как «История ярко-красной птицы», «Путешествие на Запад», «Лунный демон» и др. [4, с. 13] В них сюрреализм и таинственность придают персонажам романов сильный мифологический оттенок. В романах выделены мифологические прототипы персонажей. Главные герои онлайн-произведений, как правило, обладают аурой главного героя. Они могут родиться обычными людьми, но большинство из них одарены, способны переживать приключения, осваивать ключевые навыки, убивать демонов и монстров и отстаивать справедливость. В этом процессе главный герой часто не является зрелым, но он преодолевает различные трудности с упорством, которое обычные люди вряд ли могут достичь и осознать свою ценность.

Современные китайские онлайн-произведения тесно связаны с китайской и западной мифологией. Их авторы используют истории-прототипы, которые можно не только применять для собственного использования, пересказывать и интерпретировать их, но и полностью возрождать древние мифы в информационную эпоху.

Таким образом, исследование сетературы свидетельствует о том, что она постепенно начинает выходить за рамки развлечений и карнавала и в субкультурной форме выражает принципы и ценности господствующей идеоло-

гии китайского общества. На основе взаимодействия технологий и платформ сетература Китая ускорила темпы завоевания зарубежного рынка. Она не только открывает миру историю и культуру Китая, но и предлагает эффективную технологию создания онлайн-произведений, в том числе и в контексте мифологических образов.

ЛИТЕРАТУРА

1. 《2024年中国网络文学发展研究报告》, 2025-02-27 //中国社会科学院. = Исследовательский отчет о развитии сетературы в Китае в 2024 году, 24.09.2024 // Китайская академия общественных наук. – URL: https://www.cssn.cn/wx/tbch/202505/t20250513_5873701.shtml (дата обращения: 27.02.2025).
2. Чжэн Чжан. Литература онлайн / Чжэн Чжан // Китай. – 2018. – № 11. – С. 60–61.
3. Шань хай цзин (Каталог гор и морей) / Акад. наук СССР, Ин-т востоковедения; предисл., пер. и коммент. Э. М. Яншиной; отв. ред. Т. В. Степугина. – М. : Наука, 1977. – 235 с.
4. 胡笛.中国传统神话在网络文学中的重生.安徽文学,2017,(9).13-14. Ху, Ди. Возрождение традиционной китайской мифологии в интернет-литературе / Ху Ди // Журнал Литература. – Аньхой, 2017. – № 9. – С. 13–14.
5. Фрейд, З. Психоанализ, религия, культура : пер. с нем. / З. Фрейд. – М. : Ренессанс, 1992. – 293 с.
6. 张伟巍.中国网络玄幻小说的神话学探究.河南工程学院学报(社会科学版), 2014, (4). 69–73. Чжан, Вэйвэй. Мифологические исследования китайских интернет-фантастических романов / Чжан Вэйвэй // Журнал Хэнаньского института инженерии (издание по социальным наукам). – 2014. – № 4. – С. 69–73.

УДК 378.14

Полунина Людмила Николаевна
кандидат педагогических наук,
доцент кафедры английского языка
Тульский государственный
педагогический университет
им. Л. Н. Толстого
г. Тула, Россия

Liudmila Polunina
PhD in Education, Associate Professor
of English Language Department
Tula State Lev Tolstoy Pedagogical
University
Tula, Russia
poluninaln@mail.ru

АВТОРСТВО СТУДЕНЧЕСКОГО АКАДЕМИЧЕСКОГО ТЕКСТА: ПРОБЛЕМА ИДЕНТИФИКАЦИИ

Большие языковые модели (LLM) представляют собой значительный шаг вперед в области обработки естественного языка. Предварительное обучение на обширных корпусах текстов позволяет им демонстрировать исключительную способность генерировать высококачественный контент и универсальность при создании текстов самого разного стиля. Однако функциональные возможности больших языковых моделей порождают серьезные этические вопросы. Их способность создавать связные и контекстно-адекватные тексты повышает риск потенциального злоупотребления, включая плагиат и нарушение авторских прав. В статье представлены результаты исследования выпускных квалификационных работ студентов бакалавриата на наличие контента, сгенерированного искусственным интеллектом (ИИ). Выборку составили 20 реальных текстовых образцов в 200 слов и 20 соответствующих им по тематике и объему сгенерированных текстов. Для генерирования и анализа текстов использовались нейросеть GigaChat 2.0, а также ИИ-детекторы GigaCheck и isgen.ai. На основе сопоставительного анализа студенческих и ИИ-текстов описаны основные лингвистические особенности сгенерированного академического контента, среди которых дробная композиция, слабая когезия, отсутствие сложной научной терминологии, простой синтаксис, повторы, фейковые цитаты.

Ключевые слова: большие языковые модели; сгенерированный академический контент; ИИ-текст.

THE AUTHORSHIP OF A STUDENT ACADEMIC TEXT: AN IDENTIFICATION PROBLEM

Large language models (LLMs) represent a substantial advancement in the domain of natural language processing. Pre-trained on large text corpora, they demonstrate exceptional ability to generate high-quality content and versatility in creating texts of a wide range of styles. However, the functionality of large language models raises serious ethical questions. Their ability to generate coherent and contextually relevant texts increases the risk of potential misuse, including plagiarism and copyright infringement. This paper presents the results of a study of undergraduate students' theses for the presence of content generated by artificial intelligence (AI). The sample consisted of 20 real text excerpts of 200 words and 20 generated corresponding texts in terms of topic and volume. The texts were generated and analyzed with the GigaChat 2.0 neural network and the GigaCheck and isgen.ai AI detectors. A comparative analysis of student

and AI texts enabled revealing the main linguistic features of generated academic content, including fractional composition, weak cohesion, lack of complex scientific terms, simple syntax, repetitions, and fake quotations.

Key words: large language models; generated academic content; AI text.

В контексте стремительного развития цифровых технологий и внедрения в повседневную жизнь человека инноваций, основанных на обработке естественного языка и глубоком обучении больших языковых моделей, возникает все больше вопросов этического характера относительно авторства создаваемых вербальных текстов и произведений иных модальностей, а также относительно взаимодействия человека и искусственного интеллекта (ИИ). В исследовательском пространстве больших языковых моделей (*Large Language Models – LLMs*) и мультимодальных больших языковых моделей (*Multimodal Large Language Models – MLLMs*) сформировались два направления с диаметрально противоположными программами разработок: в то время как представители одного из них создают антропоморфные модели и обучают их максимально мимикрировать под человека при выполнении запросов и в свободном общении, маскируя маркеры искусственного интеллекта [3; 5; 7]; проponentы другого сосредоточены на совершенствовании технологий распознавания сгенерированного контента [2; 6]. Отдавая должное разработчикам инструментов, противостоящих опознанию вмешательства ИИ в процесс создания текстов, следует отметить, что большая часть их усилий направлена впоследствии на совершенствование сервисов-детекторов, которые делают видимым использование нейросетей [4].

В последние годы с проблемой «синтетического» ИИ-плагиата, который не распознается официальным программным обеспечением вузов, сталкивается все большее количество учебных заведений. Свободный доступ к различным LLM, простота их использования, не требующего специальных знаний и навыков, быстрое действие и растущее качество создаваемых ими релевантных текстовых продуктов являются основными объективными причинами массового появления сгенерированных студентами текстов – от текущих письменных заданий и конспектов до выпускных квалификационных работ. В нашем небольшом исследовании мы сосредоточимся на лингвистических особенностях реальных и сгенерированных текстов одного функционального стиля: в качестве материала для исследовательских манипуляций взяты ВКР выпускников бакалавриата по направлению подготовки «Педагогическое образование» профилей «Русский язык и Иностранный язык (английский)» и «Иностранный язык (английский)» 2021–2025 годов, посвященные методическим аспектам обучения английскому языку. В выборку вошли 20 реальных текстовых образцов примерно в 200 слов и 20 соответствующих им по тематике и объему сгенерированных текстов. Для генерирования и анализа

текстов использовались отечественная нейросеть GigaChat 2.0 (<https://giga.chat/>), ИИ-детекторы GigaCheck с общей недетализированной оценкой и isgen.ai (<https://isgen.ai/ru>), дающий процентное соотношение участия человека и ИИ в проверяемом тексте. В промптах для нейросети была дана следующая информация: тип текста – выпускная квалификационная работа (бакалаврская) по методике обучения английскому языку; объем текста – не менее 200 слов; тема в соответствии с содержанием отрывка исходного текста; краткие пояснения; необходимость включения цитат.

С целью первичной апробации регламента эксперимента был протестирован текст (218 слов) об обучении лексической стороне речи, взятый из пособия Е. Н. Солововой «Методика обучения иностранным языкам: Базовый курс лекций» [1, с. 85] и аналогичный текст (207 слов), созданный в GigaChat 2.0 по промпту: *напиши текст (200 слов) для дипломной работы бакалавра по методике обучения английскому языку о синтагматических и парадигматических отношениях на уровне лексической единицы и их значении для обучения лексической стороне речи*. Верификация реального текста в детекторах GigaCheck и isgen.ai прошла успешно с результатом 100 % вероятности того, что текст написан человеком. Тестирование сгенерированного текста показало следующий результат: GigaCheck определил его как предположительно созданный искусственным интеллектом, isgen.ai также зафиксировал участие нейросети (93,63 % общего объема текста). Экспертный сопоставительный анализ выявил следующие расхождения в структуре и вербальном оформлении текстов:

- ИИ-текст имеет общий заголовок, для него характерна композиция с делением на отдельные мелкие части, которые сопровождаются подзаголовками согласно задачам, поставленным в промпте, в отличие от реального текста, более емкого с точки зрения подачи информации;

- ИИ-текст содержит англоязычные примеры синтагматических и парадигматических отношений, в то время как Е. Н. Соловова объясняет их с учетом динамики развития синтагматических ассоциаций у младших школьников и иллюстрирует их русскими концептами;

- синтаксическое оформление ИИ-текста – выровненное, все предложения двусоставные, некоторые осложнены деепричастными оборотами и одно распространением определением, сложносочиненные и сложноподчиненные предложения в сгенерированном тексте отсутствуют в противоположность тексту учебного пособия, в котором доминируют сложные структуры, содержатся подробные пояснения в скобках, вводные и уточняющие конструкции;

- стилистически ИИ-текст выглядит более формальным и отвлеченным, тогда как в тексте Е. Н. Солововой лингвистическая информация буквально встроена в методический контекст и преломляется в соответствии с задачей

развития профессиональной компетенции будущих учителей английского языка; здесь также присутствуют оценочные суждения и вербальные маркеры более свободного стиля, такие как ... *нечего будет соединять...*, *Притом что...*, *Отсюда мы наблюдаем тенденцию...*, которые по установкам LLM не используются в академических текстах, за исключением случаев, когда нарушения стиля заявлены в промпте.

В процессе работы с выборкой текстов мы установили, что реальные текстовые образцы, взятые из проверенных в системе «Антиплагиат ВУЗ» ВКР 2020–2024 гг. с оригинальностью от 61,03 % до 85,02 %, стабильно показывали высокий процент вероятности того, что они написаны человеком, независимо от процента уникальности текста (позиции 1–5 в таблице). 2025 год стал точкой резкого роста подозрительных с точки зрения авторства работ с высокой и иногда очень высокой уникальностью (позиции 6–8), предположительно, из-за появления в свободном доступе большого количества нейросетей, обученных для генерации академических и научных текстов.

Результаты исследования авторства текстов ВКР
и текстов, сгенерированных искусственным интеллектом

	Студент	Год выпуска	Оригинальность	Оценка авторства ВКР		Оценка ИИ-текста	
				GigaCheck	isgen.ai	GigaCheck	isgen.ai
1.	Анна Л.	2021	85,02%	человек	100% человек	ИИ	100% ИИ
2.	Абдулхамид О.	2021	61,03%	человек	100% человек	ИИ	49% ИИ
3.	Екатерина Б.	2022	83,73%	человек	100% человек	ИИ	51% ИИ
4.	Камила Ч.	2023	82,11%	человек	73% человек	ИИ	76% ИИ
5.	Диана К.	2024	64,14%	человек	100% человек	ИИ	62% ИИ
6.	Юлия С.	2025	95,32%	ИИ	64% ИИ	ИИ	71% ИИ
7.	Татьяна К.	2025	58,7%	ИИ	71% ИИ	ИИ	88% ИИ
8.	Артур Х.	2025	73,03%	ИИ	53% ИИ	ИИ	57% ИИ

В связи с тем, что детекторы все еще обладают серьезными ограничениями и при определенных обстоятельствах допускают ошибки, мы провели экспертную оценку текстов, имеющих более 50 % участия нейросети, и сопоставили их с текстами, полностью сгенерированными ИИ.

В результате анализа были выявлены следующие особенности.

1. Детектор GigaCheck лучше распознает ИИ-контент родственной нейросети GigaChat, хотя следует отметить, что отсутствие детализации не позволяет оценивать это как явное преимущество.

2. Для структуры ИИ-текстов характерны короткие абзацы, буквально на одну мысль, четко определяемые пунктами промпта, а также подзаголовки, отделяющие части текста друг от друга, даже тогда, когда это нарушает когезию текста. Аналогичная композиционная особенность выявлена также в текстах ВКР, предположительно созданных с участием нейросети.

3. ИИ-тексты всегда завершаются резюме, выделяемым стандартной фразой *Таким образом, ...*. Похожая ситуация встречается в части ВКР даже там, где выводы делать не нужно.

4. ИИ-контент, как в сгенерированных, так и в студенческих текстах содержит минимальное количество сложной научной лексики и ограничен вокабуляром, не требующим обращения к дополнительным источникам информации, а также наиболее предсказуемой сочетаемостью слов. Дополнительная проверка отдельных текстов в токенизаторе ChatGPT показала незначительный процент длинных слов, превышающих пять токенов.

5. В сгенерированных текстах используются простые синтаксические структуры; минимальное количество сложноподчиненных предложений; предложения, как правило, двусоставные; не замечено сочетаний сложных предложений с деепричастными и причастными оборотами, перечислениями; перечисления выводятся в отдельные простые предложения.

6. ИИ-тексты отличаются слабой когезией за счет распределения контента четко по задачам промпта.

7. В сгенерированных текстах ВКР часто встречается многократное повторение одного и того же пассажа, перефразируемого на разные лады, что делает текст практически нечитаемым. С нашей точки зрения, такая тактика нейросети объясняется недостаточно четкой и подробной формулировкой промпта, что заставляет модель буквально «бегать по кругу», набирая необходимое количество слов.

8. Для текстов ВКР с участием ИИ характерно полное отсутствие цитат в сочетании с внушительным списком литературы и высоким процентом уникальности, достигающим при проверке в системе Антиплагиат ВУЗ 95 % и более. Это наиболее очевидный признак сгенерированного текста, поскольку даже произвольно расставленные постфактум ссылки принципиально не могут исправить ситуацию, более того, недостоверное цитирование легко идентифицируется при экспертном разборе текста ВКР.

9. При наличии в промпте задачи включить цитаты в текст нейросеть замечена в генерации фейковых цитат и придумывании несуществующих ученых, как правило, с самыми распространенными именами и фамилиями. В качестве примеров приведем следующие «цитаты» (фамилии сокращены):

(1) *«Иностранный язык должен восприниматься детьми не как абстрактная дисциплина, а как инструмент общения», – подчеркивает доктор педагогических наук Ирина Петровна С.*

(2) *«Для успешного изучения языка важно создание условий погружения в языковую среду», – утверждает кандидат филологических наук Ольга Владимировна И.*

Обе цитаты находятся в одном и том же тексте, через небольшой абзац друг от друга. Их содержание в стиле Captain Obvious, безусловно, соответствует тематике ВКР по методике обучения любому иностранному языку, как и любому его аспекту или виду речевой деятельности. В первом случае ни одна поисковая система, включая саму нейросеть, не смогла обнаружить доктора педагогических наук с такими данными, несмотря на услужливо предложенную моделью расшифровку имени и отчества. Во втором случае цитата приписана реальному человеку, но тот не имеет никакого отношения к иностранным языкам, занимаясь совершенно другим научным направлением.

Следует отметить, что любая нейросеть наряду с генерацией текста предлагает его «очеловечить» (*humanize*), перефразировать (*paraphrase*) или переписать (*rewrite*), чтобы избавиться от маркеров искусственного интеллекта. Полностью русскоязычные нейросети, обучаемые изначально на русском языке, пока совсем немногочисленны, равно как и другие сервисы для работы с текстом и другими модальностями. Использование англоязычных и мультилингвальных моделей при работе с текстами на русском языке не дает достаточной гарантии точности или даже истинности результата.

Для сравнения мы опробовали описанный выше исследовательский регламент на англоязычном тексте, созданным в ChatGPT, по промпту: *write a short paragraph of 170 to 200 words for a bachelor thesis on teaching English as a foreign language about vizualization as a means of semantization of vocabulary, include quotes of well-known scientists*. Все доступные англоязычные детекторы с результатом 100 % определили его как ИИ-текст, этот результат не смогло изменить дальнейшее перефразирование, выполненное в нескольких нейросетях. Среди достоинств ChatGPT можно отметить реальные, пусть и слегка устаревшие, ссылки на работы действительно известных ученых, грамотную компоновку информации, использование типичной для методических статей терминологии.

Получается, что преподаватели и администрация вузов имеют возможность распознавать неправомерное использование нейросетей студентами.

И университеты уже сейчас вводят в уставы ограничения, связанные с использованием систем искусственного интеллекта. Однако решение такой сложной проблемы не может быть простым, оно требует продвижения в нескольких направлениях. Безусловно, это и технологическое совершенствование LLM в направлении интегрирования этических стандартов и стандартов безопасности. Не меньшее значение имеет также сочетание потенциала больших языковых моделей и экспертной оценки человека, так как человеческая интуиция, опыт и чувство языка зачастую превосходят по результативности статистические возможности нейросетей. Но самым важным проектом все же должна стать работа со студентами, для того чтобы они понимали, что генерировать действительно новые, значимые и креативные идеи может только человек. И этому нужно учиться самостоятельно, без подсказок «умных» машин.

ЛИТЕРАТУРА

1. Соловова Е. Н. Методика обучения иностранным языкам : Базовый курс лекций : Пособие для студентов пед. вузов и учителей. М. : Просвещение, 2002. 239 с.
2. A large-scale comparison of human-written versus ChatGPT-generated essays [Electronic resource] / Herbold S., Hautli-Janisz A., Heuer U., Kikteva Z., Trautsch A. // Scientific Reports. 2023. N. 13. URL: <https://www.nature.com/articles/s41598-023-45644-9> (date of access: 26.06.2025). DOI: <https://doi.org/10.1038/s41598-023-45644-9>.
3. Can AI-Generated Text be Reliably Detected? Stress Testing AI Text Detectors Under Various Attacks [Electronic resource] / Sadasivan V. S., Kumar A., Balasubramanian S., Wang W., Feizi S. // Transactions on Machine Learning Research. 2025. N. 1. URL: <https://openreview.net/pdf?id=OOgsAZdFOt> (date of access: 23.06.2025).
4. Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense [Electronic resource] / Krishna K., Song Y., Karpinska M., Wieting J., Iyyer M. // Proceedings of the 37th International Conference on Neural Information Processing Systems, 2023. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/575c450013d0e99e4b0ecf82bd1afaa4-Paper-Conference.pdf (date of access: 26.06.2025).
5. Simple techniques to bypass GenAI text detectors: implications for inclusive education [Electronic resource] / Perkins M., Roe J., Vu B. H., Postma D., Hickerson D., McGaughran J., Khuat H. Q. // International Journal of Educational Technology in Higher Education. 2024. N. 21. URL: <https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-024-00487-w#Bib1> (date of access: 26.06.2025) DOI: <https://doi.org/10.1186/s41239-024-00487-w>.

6. SKDU at De-Factify 4.0: Natural language features for AI-Generated Text-Detection [Electronic resource] / Maviya Sh., Arnau-González P., Arevalillo-Herráez M., Katsigiannis S. // 39th Annual AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA. 2025. URL: <https://arxiv.org/abs/2503.22338> (date of access: 01.07.2025).
7. Ying Zhou, Ben He, Le Sun. Humanizing Machine-Generated Content: Evading AI-Text Detection through Adversarial Attack [Electronic resource] // The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. 2024. P. 8427–8437. URL: <https://aclanthology.org/2024.lrec-main.739.pdf> (date of access: 26.06.2025).

УДК 811.161.1

Сафарли Лейла Шамиль гызы
доктор философии по филологии,
преподаватель
Бакинский славянский университет
г. Баку, Азербайджан

Leyla Shamil Safarli
PhD in Philology, Lecturer
Baku Slavic University
Baku, Azerbaijan
leylasafarli@mail.ru

ЛИНГВОКУЛЬТУРОЛОГИЯ МЕДИАТЕКСТА: ТЕОРЕТИЧЕСКИЕ ОСНОВАНИЯ И АНАЛИТИЧЕСКИЕ ПЕРСПЕКТИВЫ

В статье рассматриваются теоретические основания и аналитические перспективы нового научного направления — лингвокультурологии медиатекста, сформировавшегося на стыке лингвистики, культурологии и медиакоммуникации. Обосновывается необходимость изучения медиатекста как особого лингвокультурного феномена, обладающего высокой семиотической и социокультурной значимостью. Раскрываются ключевые понятия и методологические подходы лингвокультурологического анализа медиаречи, включая концептуальный, сопоставительный и филологический методы. Отдельное внимание уделяется роли медиатекстов в трансляции культурных кодов, формировании языковой личности и репрезентации ценностных ориентиров общества. Подчеркивается значимость изучения неологизмов, прецедентных феноменов и культуроспецифической лексики как индикаторов социокультурных процессов. В статье анализируются идеи В. А. Богородицкого о социокультурной природе языка и ассоциативной основе значения слова, которые актуализируются в рамках современного медиaprостранства. Делается вывод о необходимости комплексного подхода к исследованию медиатекста как средства отражения и конструирования современной языковой и культурной картины мира.

Ключевые слова: лингвокультурология; медиатекст; массовая коммуникация; культурнозначимая информация; языковая личность; культуроспецифическая лексика.

LINGUOCULTUROLOGY OF MEDIA TEXT: THEORETICAL FOUNDATIONS AND ANALYTICAL PERSPECTIVES

The article explores the theoretical foundations and analytical prospects of the emerging interdisciplinary field — linguoculturology of media texts — which integrates linguistics, cultural studies, and media communication. The necessity of studying media texts as complex linguocultural phenomena with high semiotic and sociocultural relevance is substantiated. Key concepts and methodological approaches to the linguoculturological analysis of media discourse are outlined, including conceptual, comparative, and philological methods. Special attention is given to the role of media texts in transmitting cultural codes, shaping linguistic identity, and representing societal value orientations. The article emphasizes the importance of studying neologisms, precedent phenomena, and culturally specific vocabulary as indicators of sociocultural processes. The ideas of V. A. Bogoroditsky on the sociocultural nature of language and the associative basis of word meaning are analyzed in the context of modern media. The paper concludes by underscoring the need for a comprehensive approach to examining media texts as instruments for reflecting and constructing the contemporary linguistic and cultural worldview.

Key words: linguoculturology; media text; mass communication; culturally significant information; linguistic personality; culture-specific lexicon.

В начале XXI века лингвокультурология становится одной из ключевых областей современного языкознания, демонстрируя высокую исследовательскую активность и перспективность. Она не ограничивается лишь утверждением устойчивой связи между языком и культурой, а также стремится выявить и объяснить закономерности влияния социокультурных процессов на формирование языковой личности и развитие лингвистической культуры общества.

В современном научном дискурсе все большее внимание уделяется исследованию языковых особенностей средств массовой информации. Это связано с тем, что СМИ играют ключевую роль в формировании общественного мнения и мировоззренческих установок. Они прочно интегрированы в общественную и коммуникативную жизнь, активно взаимодействуют с различными сферами социума, включая литературу, и оказывают заметное влияние на развитие и функционирование литературного языка. При анализе языковых изменений под воздействием СМИ важно учитывать не только влияние цифровых технологий, но и более глубокие трансформации, происходящие в лингвокультурной среде. Как подчеркивает Т. Г. Добросклонская, речь идет прежде всего о качественных изменениях в языке, отражающих изменение всей системы культурной и речевой коммуникации [4, с. 11].

Термин «медiateкст», широко используемый в современных научных исследованиях, охватывает широкий спектр текстов, принадлежащих различным медийным форматам и жанрам: от рекламных и PR-сообщений до журналистских материалов и интернет-контента. Современный медiateкст представляет собой сложное лингвокультурное образование, находящееся на пересечении интересов таких гуманитарных дисциплин, как семиотика, философия, психология. Тем не менее, лингвокультурологический подход, по нашему мнению, продолжает занимать центральное место в изучении этого феномена, поскольку именно медiateксты активно участвуют в формировании языковой личности.

Следует отметить, что научным сообществом уже давно признано: язык современности активно обновляется именно через сферу средств массовой информации. При этом медиаречь, не имея тематических границ, включает в себя и культурные элементы. В этом контексте особенно важен взгляд на язык как на компонент культуры, способный не только фиксировать и накапливать культурно значимую информацию, но и активно влиять на трансформации культурного пространства. Такой подход открывает широкие возможности для анализа медiateкстов. Как подчеркивает Т. Г. Добросклонская, «взгляд на язык как фактор культуры, обладающий способностью отражать, сохранять и накапливать культурозначимую информацию, а также влиять на динамику культурного контекста, имеет огромное значение для анализа текстов массовой информации» [4, с. 221].

В этой связи особую значимость приобретает культурологический подход к интерпретации текстов массовой коммуникации. Он позволяет не

только выявить культурно маркированные элементы, но и проследить как общие, так и уникальные черты национальных культур и глобальных культурных процессов.

В рамках развития современной гуманитарной мысли можно говорить о становлении нового научного направления – лингвокультурологии медиатекста. Объектом данной дисциплины выступает медиатекст как сложное коммуникативное явление, функционирующее в пространстве массовой информации и коммуникации. Предметом исследования становятся языковые особенности медиатекстов, через которые транслируются культурные коды и осуществляется влияние на развитие культурной парадигмы общества.

Подобно другим новым направлениям, лингвокультурология медиатекста стремится обрести теоретические основания, обращаясь к трудам выдающихся лингвистов. Среди них особое место занимает В. А. Богородицкий, чьи идеи о социокультурной природе языка, механизмах языковых изменений, психологических аспектах ассоциативного мышления и роли языка в обществе остаются актуальными и сегодня.

В своей фундаментальной работе «Лекции по общему языкознанию» В. А. Богородицкий детально обозначил основные векторы лингвистического исследования, рассматривая язык как динамическое, эволюционирующее явление. Особенно ценными для современных лингвокультурологических исследований медиатекста являются его идеи о связи языка с социокультурными процессами. Ученый справедливо указывал на неразрывную связь между развитием языка и прогрессом цивилизации: «...прогресс языка тесно связан с прогрессом цивилизации, так как появление у человека новых идей и понятий неизменно сопровождается появлением новых слов и выражений; таким образом, язык является как бы летописью пережитой культурной и социальной истории народа» [2, с. 7].

С этих позиций можно рассматривать феномен неологизации, особенно актуальный в контексте медиакommunikации. Медиатексты не только фиксируют, но и активно продуцируют неологизмы, многие из которых отражают социально-культурные трансформации. Возникновение новых слов в медиaprостранстве тесно связано с активностью интернет-пользователей, в рамках концепции гражданской журналистики. Любой пользователь, публикующий сообщение в сети, сегодня может претендовать на роль автора новостного события или интерпретатора происходящего. При этом нередки случаи, когда такие «авторы» не обладают профессиональной языковой подготовкой, однако именно они становятся источниками новых словоформ. Так, в журналистскую и медийную лексику вошли неологизмы интернет-пользователей – например, «Луганда», «вата», «сетевые хомячки». Эти языковые новообразования не только иллюстрируют творческий потенциал массового пользователя, но и дают представление о культурных доминантах современного медиaprостранства, среди которых нередко – агрессивность, поляризация взглядов и отказ от конструктивного диалога.

Одним из важнейших аспектов понимания природы медиатекста является учение В. А. Богородицкого о психологических основах языковой ассоциативности. Он подчеркивал, что значение слова формируется на основе комплекса ассоциаций, опирающихся на чувственные восприятия внешнего мира: «...само значение слова, как и звуковое его выражение, представляет собой сложный агрегат ассоциаций, складывающийся из множества элементарных ощущений» [2, с. 104].

Ассоциативность – ключевая характеристика медиатекста, обуславливающая его воздействующую силу. Медиатексты, как правило, кратковременны, но чрезвычайно насыщены прагматическим содержанием, направленным на формирование определенного поведения целевой аудитории. Именно поэтому в них столь востребованы культурные прецедентные феномены – устойчивые выражения, образы, метафоры, апеллирующие к общекультурной памяти и вызывающие у читателя/зрителя мгновенные ассоциации. Такие элементы выполняют роль культурных маркеров, актуализируют национальные и универсальные культурные стереотипы, способствуют осмыслению информации и формированию отношения к социальным реалиям.

Настоящее исследование, выполненное в русле лингвокультурологического подхода, акцентирует внимание на анализе языковой картины мира, отраженной в медиатекстах. На наш взгляд, именно СМИ становятся важнейшим инструментом формирования современной картины мира, поскольку они представляют разнообразие реальности в разных форматах – от печатных до цифровых.

В этом контексте уместно рассматривать взаимосвязь таких понятий, как «языковая картина мира», «публицистическая картина мира» и «информационная картина мира». Публицистическая картина мира формируется средствами языка, транслирует авторскую точку зрения, воздействует на читателя и одновременно влияет на нормы литературного языка. Информационная картина мира, в свою очередь, репрезентирует культурно-ценностные особенности мышления и восприятия, отражая менталитет и ценностные ориентиры той или иной нации.

Лингвокультурологический подход приобретает особую значимость в контексте анализа текстов массовой коммуникации, которые оказывают существенное влияние на развитие современного русского языка. В условиях информационного общества закономерным становится обращение к понятию лингвокультурологических процессов, отражающих динамику взаимодействия языка и культуры в медиапространстве.

Так, например, Т. Г. Добросклонская подчеркивает: "Лингвокультурология свидетельствует о становлении новой антропологической парадигмы в гуманитарной науке. Значение лингвокультурологического метода для анализа текстов массовой информации очевидно, поскольку весь корпус такой

речи является важнейшим компонентом со временной культуры. Медиа-тексты насыщены культурозначимой информацией, в них фиксируются и отражаются как общие, так и специфические особенности функционирования национальных языков и национальных культур. Именно поэтому выявление и описание культурозначимых компонентов медиаречи оказывается одной из важнейших исследовательских задач" [5, с. 45].

Задачи лингвокультурологического анализа включают [3, с. 3]:

- 1) способность извлекать информацию о культурных особенностях страны из текста;
- 2) умение интерпретировать и анализировать полученные данные;
- 3) способность проводить сравнительный анализ ценностных картин мира различных культур.

Методы лингвокультурологического анализа [3, с. 3]:

- метод филологического анализа – включает приемы жанровой интерпретации языковых средств и анализа идейного содержания текста;
- метод концептуального анализа – направлен на реконструкцию картины мира говорящего;
- сопоставительный метод – используется для сравнения культурных и языковых особенностей различных языков и их носителей [6, с. 3].

Предмет лингвокультурологического анализа включает языковые единицы, которые приобрели символическое, эталонное и образно-метафорическое значение в культуре, обобщая результаты человеческого сознания – как архетипические, так и прототипические. Эти единицы фиксируются в мифах, легендах, ритуалах, обрядах, фольклоре, религиозных дискурсах, поэтических и прозаических текстах, фразеологизмах, метафорах, символах и поговорках (поговорках). Они отражают национально-культурную специфику и аккумулируют культурный опыт, передаваемый из поколения в поколение.

К таким единицам относятся [5, с. 5]:

- Безэквивалентная лексика и лакуны
- Мифологизированные языковые единицы: архетипы и мифологемы
- Эталоны, стереотипы, символы
- Паремии
- Фразеологизмы
- Метафоры и образы языка
- Речевое поведение
- Речевой этикет (по В. А. Масловой) [7]:
- Ключевые концепты культуры
- Прецедентные феномены

При анализе культурообусловленных элементов текста речь идет о широком спектре языковых единиц: от слов, обозначающих специфические

реалии (таких как *Кремль*, *Белый дом*, *Биг Бен*, *Би-Би-Си*), культуроспецифической лексики и безэквивалентных слов (например, *дача*, *гласность*, *челноки*) до заимствованных терминов и выражений (например, *fin de siècle* ‘конец века’, *haute couture* ‘высокая мода’ и другие). Важно подчеркнуть практическое значение лингвокультурологического анализа, так как его целью является выявление и интерпретация культурозначимых единиц, которые играют ключевую роль при составлении культурологического комментария.

Аналитические перспективы лингвокультурологии медиатекста проявляются через несколько ключевых направлений, нацеленных на исследование взаимодействия языка и культуры в рамках современных медиатекстов. Существенно отметить, что медиатексты, являясь культурными и коммуникативными феноменами, подвергаются разнообразным трансформациям, отражающим динамику изменений в социальных, политических и культурных аспектах.

Ниже представлены основные аналитические направления.

❖ Изучение культурных кодов в медиатекстах

Медиатексты представляют собой пространство, в котором активно передаются культурные коды, имеющие как универсальные, так и специфические характеристики для конкретной культуры. Исследование этих кодов позволяет определить, каким образом определенные концепты, символы, стереотипы и мифы воздействуют на восприятие информации аудиторией.

❖ Анализ неологизмов и изменений в языке

Средства массовой информации играют ключевую роль в возникновении неологизмов и языковых инноваций, которые отражают изменения в социокультурной реальности. Изучение неологизмов помогает проследить, как язык медиа приспосабливается к культурным и технологическим трансформациям, а также как массовое внедрение новых слов в медиатексты сказывается на эволюции языка.

❖ Роль медиатекстов в формировании языковой личности

Медиатексты оказывают влияние на языковую личность, формируя новые модели речевого поведения, предпочтений и даже мировоззрения. Исследование этого влияния помогает понять, как массовая коммуникация способствует определению норм и стандартов языка в обществе.

❖ Влияние медиатекстов на культурную картину мира

Медиатексты активно участвуют в формировании культурной картины мира, транслируя мировоззренческие установки и ориентиры. Анализ этих текстов позволяет понять, как средства массовой информации отражают и одновременно формируют восприятие культурных и социальных процессов.

❖ Анализ ассоциативности и психолингвистических аспектов медиатекстов

Психолингвистические исследования, основанные на идеях В. А. Богородицкого, акцентируют внимание на ассоциативных процессах, которые возникают у аудитории при восприятии медиатекстов. Это включает изучение, как медиатексты активируют культурные и личные ассоциации, влияя на восприятие информации.

❖ Сравнительный анализ медиатекстов разных культур

Важным аспектом аналитической перспективы является использование сравнительного метода для исследования медиатекстов в разных культурных контекстах. Это помогает выявить универсальные и специфические черты в построении медиатекстов и их восприятии.

❖ Методы лингвокультурологического анализа

Использование различных методов анализа, таких как концептуальный, сопоставительный и филологический, позволяет глубже понять, как медиатексты служат инструментами культурной и языковой коммуникации, а также как они отражают изменения в обществе и языке.

❖ Критическое восприятие медиатекстов

В итоге, критический анализ медиатекстов с точки зрения их воздействия на общественное сознание и мировоззрение становится важной перспективой. Это включает в себя оценку того, как СМИ могут манипулировать языком и культурой в целях воздействия на массовую аудиторию.

Эти перспективы открывают широкие возможности для дальнейших исследований в области лингвокультурологии медиатекста и позволяют глубже понять роль языка и культуры в современных средствах массовой информации.

Практическая значимость данного исследования заключается в возможности применения лингвокультурологического подхода для анализа языковых изменений, происходящих в современных медиатекстах. Выявление культурно маркированных единиц в медиатекстах способствует лучшему пониманию культурных и социальных процессов, происходящих в обществе, и помогает разработать более эффективные способы работы с текстами массовой информации, как в научных исследованиях, так и в сфере PR, рекламы и журналистики.

Лингвокультурология медиатекста является перспективным направлением, позволяющим глубже понять механизмы влияния языка на общественные и культурные процессы. Медиатексты, как сложные лингвокультурные образования, играют ключевую роль в формировании не только языковой личности, но и в трансляции культурных ценностей. Особое внимание стоит уделить тому, как медиасреда активно создает и распространяет неологизмы, метафоры и прецедентные феномены, которые становятся основой для формирования общественного мнения и культурных стереотипов. Лингвокульту-

рологический анализ медиатекстов помогает выявить эти процессы и осознать их влияние на развитие языка и культуры в условиях глобализации и технологических изменений.

ЛИТЕРАТУРА

1. Ахундова М. Б. Лингвокультурологические тексты как средство формирования коммуникативной компетенции учащихся. *Pedaqogika, psixologiya elmləri üzrə elmi-nəzəri-metodik jurnal*. № 3, 2021. С. 131–136.
2. Богородицкий В. А. Лекции по общему языкознанию / В. А. Богородицкий. – Казань, 1911. – 246 с.
3. Ворошкевич Д. В., Казанникова Д. П. Пособие по лингвокультурологическому анализу текста. М-во образования и науки Российской Федерации, Федеральное гос. бюджетное образовательное учреждение высш. проф. образования "Московский пед. гос. ун-т". – М. : МПГУ, 2016. – 39 с.
4. Добросклонская Т. Г. Медиалингвистика: системный подход к изучению языка СМИ (современная английская медиаречь): учебное пособие / Т. Г. Добросклонская. – М. : Флинта, Наука, 2008. – 263 с.
5. Добросклонская Т. Г. Язык средств массовой информации / Т. Г. Добросклонская. – М. : КДУ, 2015. – 116 с.
6. Игошина Оксана Алексеевна. Лингвокультурологический анализ фольклорного текста при обучении русскому языку немецкоязычных студентов: Дис. ... канд. пед. наук: 13.00.02: Санкт-Петербург, 2003. 270 с. РГБ ОД, 61:04-13/818
7. Маслова В. А. М 31 Лингвокультурология: Учеб. пособие для студ. Высш. Учеб. заведений. – М. : Издательский центр «Академия», 2001. – 208 с.
8. Фаткуллина, Ф. Г. Лингвокультурологический подход к изучению текстов массмедиа / Ф. Г. Фаткуллина, Л. М. Салимова, Д. К. Саяхова // Казанская наука. – 2019. – № 10. – С. 61–63.

УДК 811'111

Свистун Татьяна Ивановна

кандидат филологических наук,
заведующий кафедрой теории
и практики английского языка
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Tatiana Svistun

PhD in Philology,
Head of the Department
of Theory and Practice of English
Belarusian State University
of Foreign Languages
Minsk, Belarus
tess17@mail.ru

Русанов Алексей Александрович

выпускник
Белорусский государственный
университет иностранных языков
г. Минск, Беларусь

Aleksei Rusanov

University Graduate
Belarusian State University
of Foreign Languages
Minsk, Belarus
secretjunior111@gmail.com

СТРУКТУРНО-ЯЗЫКОВАЯ СПЕЦИФИКА СТАТЬИ-ЛИСТИКЛА

Интернет-дискурс быстро развивается. Используемые в нем технологии позволяют создавать новые формы взаимодействия с целевой аудиторией, одной из которых является жанр «листикл», представляющий собой текст в виде нумерованного или маркированного списка. Для рассматриваемого жанра характерно активное использование мультимедиа и гиперссылок. Особенностью листикла как жанра интернет-дискурса также является сочетание разговорных выражений и литературной речи, что делает тексты статей более доступными и понятными пользователям. Типичная структура листикла состоит из заголовка, вводной части, основной части и заключения. Анализ листиклов в сфере отдыха и туризма на русском и английском языках позволил выявить языковые особенности с точки зрения лексики, стилистики и синтаксиса: числительные в заголовках, оценочные прилагательные и топонимы, географические термины, сравнения и метафоры при описании туристических объектов и услуг в основной части, побудительные предложения и риторические вопросы. Англоязычные и русскоязычные листиклы имеют больше сходств, чем различий. Обнаруженные сходства могут быть обусловлены общими тенденциями в сфере SEO-оптимизации, интернет-журналистики и маркетинга, где ключевыми задачами являются повышение читаемости текстов и вовлеченности аудитории. Стандарты, задаваемые глобальной интернет-средой, способствуют унификации форматов контента, что в значительной степени влияет на структуру и стиль листиклов, независимо от языка написания.

Ключевые слова: интернет-дискурс; листикл; нумерованный/маркированный список; мультимедиа; гиперссылка; языковые средства.

STRUCTURAL AND LINGUISTIC PECULIARITIES OF A LISTICLE

Internet-discourse is developing quickly. Its technologies make it possible to create new forms of interaction with target audience, for example the genre of a listicle, which is a text structured as numbered or marked lists as the name suggests. The genre under study is characterized by active usage of multimedia and hyperlinks. Another peculiar feature is the

combination of conversational and literary expressions, which makes the texts of the articles easy to perceive and understand. The typical structure of a listicle includes the headline, lead-in, main part and conclusion. The analysis of listicles in the sphere of recreation and tourism helped to reveal the linguistic means from the point of lexis, stylistics and syntax: numerals in headlines, evaluative adjectives and toponyms, geographical terms, comparisons and metaphors in descriptions of tourist destinations and services in the main part, imperative sentences and rhetorical questions. English and Russian listicles have more similar features than differences, which might be determined by common tendencies in SEO-optimization, internet-journalism and marketing where the key aims are to increase the readability of the texts and users' engagement. Global net standards require the unification of content formats, which largely influences the structure and the style of listicles irrespective of the language.

Key words: internet-discourse; listicle; numbered/marked list; multimedia; hyperlink; language means.

В современном информационном обществе Интернет становится все более значимым и влиятельным средством коммуникации. С развитием онлайн-платформ и социальных сетей возникают новые жанры, формы общения и передачи информации, терминологическая вариативность и количественная неоднозначность которых вызывает много вопросов [1, с. 47] и требует особого внимания лингвистов. Среди новых жанров следует выделить «листикл» (от англ. *list* + *article*) – статьи, представленные в виде списка, в краткой форме описывающего то или иное явление. Такая форма подачи материала отвечает требованиям информационной эпохи и удовлетворяет необходимость передачи значительного объема информации в сжатом виде, что привлекает множество создателей интернет-контента, в том числе и коммерческого. Такие статьи также представляют особый интерес для лингвистов с точки зрения особенностей жанровой и языковой специфики.

Актуальность данной темы обусловлена тем, что листикл как жанр, использующийся в интернет-журналистике и маркетинге, является популярной, привлекающей трафик и легко понимаемой формой. Листикл помогает организовывать информацию и делать ее более доступной. Выделяя листикл как отдельный жанр интернет-дискурса, стоит отметить, что некоторые исследователи используют различные термины для его обозначения. В. В. Деменьтев, например, использует термин *интернет-рейтинг* [2, с. 230], а В. В. Фролова называет рассматриваемый жанр *статьей-списком*, или *листиклом* [3, с. 160]. Согласно мнению В. В. Фроловой, статью-список можно определить как «относительно новый жанр информационно-развлекательной журналистики, использующий в качестве композиционного принципа организации текста нумерованный список, что почти обязательно отражается в заголовке» [3, с. 162–163]. В англоязычной литературе распространен термин *листикл* (от англ. *listicle*).

В качестве материала исследования были отобраны 200 статей-листиклов в сфере отдыха и туризма на русском и английском языках (по 100 статей на каждом языке соответственно). Корпус исследования был проанализиро-

ван с точки зрения структуры для выделения обязательных и факультативных элементов статей. Кроме того, были выявлены лексико-стилистические и синтаксические средства, присущие рассматриваемому жанру интернет-дискурса.

Типичная структура листикла включает заголовок, вводную часть (lead-in), основную часть и заключение. Обязательными являются заголовок, вводная часть, основная часть, мультимедиа (изображения, видеозаписи, gif-анимации, интерактивные объекты). Факультативными элементами являются подзаголовок, гиперссылки, оглавление, а также заключение. Заголовок является самым главным элементом листикла. Он имеет кликбейтный характер и, соответственно, содержит языковые и структурные элементы, привлекающие наибольшее внимание пользователей. Вся суть статьи отражается в заголовке, его построение продиктовано нуждами поисковой, или SEO, оптимизации (от англ. *Search Engine Optimization*). Вводная часть, в свою очередь, также является обязательным элементом листикла. Длина данного элемента может варьироваться от одного-двух предложений, и его наличие в таком случае скорее формально, до нескольких абзацев, подробно раскрывающих тему. Далее, перед основной частью факультативно может присутствовать оглавление, которое значительно упрощает навигацию пользователя, что позитивно сказывается на вовлечении. Основная часть листикла в подавляющем большинстве случаев представляет собой нумерованный список, порядок в котором может быть обусловлен важностью или значимостью элементов, но элементы списка могут быть расположены и в произвольном порядке. Элементы ранжируются как по популярности, так и по личным соображениям автора статьи. В некоторых случаях список может быть ненумерованным, а отдельные сайты предлагают после прочтения заголовка и введения интерактивную доску, на которой располагаются страны, отели, или достопримечательности, кликнув на которую читатель попадает в один из пунктов списка. Длина основной части может включать от нескольких предложений до 2–3 абзацев. Сами элементы списка также разбиты на блоки с обязательными внутритекстовыми заголовками и иллюстрациями.

Пункты списка обязательно сопровождаются различными мультимедиа, как правило это фотографии и видео, как из сторонних источников, так и непосредственно от авторов статьи. Как и для других жанров интернет-дискурса, для жанра листикл характерна полимодальность. В целом, включение в интернет-тексты буквенных, образно-зрительных и образно-слуховых компонентов в виде изображений, аудио- и видеовставок привлекает внимание пользователя сети и стимулирует интерпретацию с его стороны [4, с. 97].

Широко фиксируется наличие в листиклах гиперссылок на статьи по теме, которые встраиваются в основной текст либо размещаются в виде отдельных блоков в конце или начале текста. Включение гиперссылок в листиклы имеет несколько важных функций. Гиперссылка позволяет получить дополнительную информацию при переходе к другому интернет-

ресурсу, улучшить навигацию по сайту и структурировать контент, связывая воедино статьи по смежной тематике, повысить SEO-показатели статьи за счет внутренних перекрестных ссылок.

Факультативным элементом листика является заключение. Ввиду концентрированного и емкого характера подачи информации в подобных статьях, заключение – это крайне редко встречающийся структурный элемент, который представляет собой краткое резюме (не более нескольких предложений). Помимо этого, пользователи редко читают всю статью полностью или дочитывают до конца, вследствие этого данный структурный элемент не играет значительной роли в листиках.

Рассмотрим языковые (лексические, стилистические, синтаксические) особенности рассматриваемого интернет-жанра. По результатам анализа, самым распространенным средством, зафиксированным в подавляющем большинстве листиков, в частности в заголовках, стали числительные, чаще кратные 5. Приведем некоторые примеры, сохраняя орфографию и пунктуацию оригиналов здесь и далее: **25** *places to visit before you turn 25*; **10** *Best Places to Travel on a Budget*; **19** *Best Places To Visit In Morocco In 2024 That Will Take Your Breath Away!*; **15** *мест, куда съездить на выходные в Беларуси*; **10** *самых интересных мест для путешествий*; **100** *самых красивых мест на планете*; **Top-38**. *Главные достопримечательности Дубая*.

Не менее широко распространено употребление оценочных прилагательных, например *top*, *marvelous*, *pretty*, *красивый*, *удивительный*, *необычный*, *интересный* особенно в превосходной степени: **Top** *tourist attractions & places to visit in Scotland*; **20 Prettiest And Best Places To Visit In England**; **28 Landscapes So Marvelous That It Will Make Your Trip Worth It!**; **World's most remote**, *yet rewarding, places to visit*; **40 самых красивых мест столицы**; **ТОП-10 самых необычных мест мира, словно с другой планеты**; **Достопримечательности Англии: ТОП-25 самых интересных мест в 2025**; **50 удивительных** и **красивейших мест на Земле**; **Вы могли не знать: 25 необычных мест Беларуси**. Оценочные прилагательные указывают либо на характерную или уникальную черту объекта, либо на мнение автора.

Неотъемлемым компонентом листика в сфере туризма и отдыха является наличие топонима (регион, страна, город): **15 Best Places to Visit in Belarus – The Crazy Tourist**; **Bucket List Travel: The Top 20 Places In The World**; **50 beautiful places to visit in South Korea**; **Top 10 places to visit in Vienna**; **25 Best Places to Visit in Argentina (Chosen By Experts!)**; **28 Best Tourist Places in Indonesia in 2025**; **15 мест, куда съездить на выходные в Беларуси**; **Достопримечательности Минска – ТОП-25 самых популярных мест**. Спектр топонимов достаточно разнообразен. Были зафиксированы как объекты мирового масштаба (континенты, страны, города), так и локальные (малые населенные пункты, районы).

Наблюдается активное использование разговорной лексики, что делает тексты более доступными и понятными рядовому пользователю. Сюда относятся неполные предложения, фразовые глаголы (в английском языке), жаргонизмы: **Can't decide?** *Travel the island's entire 110-mile coastline...*;

*Planning a holiday to Europe?; Looking to **feed off** the energy of two ecstatic fan bases? How many can you **tick off**?; **Отдельный большой респект** за ресторанный Инстаграм на белорусском языке, в котором не только **аппетитные фото** и описания кулинарных шедевров от Чикилевского... Как известно, многие интернет-жанры, в том числе и институциональные, тяготеют к разговорной речи, сленгу или жаргону [5, с. 351]. И жанр листикл не стал исключением.*

Особое место среди выявленных языковых средств занимают стилистические средства. Использование в листиклах в сфере туризма и отдыха сравнений, эпитетов и метафор, описывающих туристические объекты и услуги, позволяет более ярко представить тему. Существительные *heart, jewel, gem, paradise* в англоязычных статьях-листиклах и *сердце, жемчужина, рай, душа* – в русскоязычных выступают как примеры в данном случае: *Sardinia is a **jewel** in the Mediterranean crown. Stretching from the **heart** of the Apennines to the Adriatic Sea...; **unspoiled villages, picturesque** Trabocchi Coast, and **stunning** natural escapes; Старый Тбилиси – это **сердце** города, место, где современность органично дополняет историю, отражая дух грузинской столицы; Телецкое озеро – **жемчужина** Алтая; Вьетнам – это **рай на Земле**; Россия – не просто **щедрая душа**, но ещё и тысячи мест, где можно этой самой душой отдохнуть.*

Кроме того, для создания яркого и запоминающегося образа места или города часто используются глаголы чувственного восприятия, затрагивающие зрение, слух, осязание. Рассмотрим следующие примеры на обоих языках: *But for aspiring naturalists, the journey is well worth it for the opportunity to **see** wildlife that can't be found anywhere else on Earth; **Listen** out for the firing of the One O'Clock Gun from the castle grounds, which **can be heard** daily; С нами, кстати, вы можете **не только увидеть, но и услышать** Париж – загляните в наш магазин аудиогидов; В солнечную погоду **насладитесь** напитком на центральной площади, и вы будете **чувствовать себя так, как будто оказались в раю**.*

Также, текст основной части листиклов характеризуется активным использованием различных туристических клише, к примеру – *hidden gems, off the beaten path, must-see, a picture postcard view, a true icon, a slice of paradise*, рассмотрим их использование на примерах: *When you visit Italy's capital city, prepare to cross a few **must-see landmarks** – including the Colosseum, the Trevi Fountain and the Pantheon – **off of your bucket list**; Cappadocia is home to **gems of rare beauty** such as the Goreme National Park; A sign now marks the spot beside Durmitor's **postcard** attraction, the Black Lake; When it comes to wine regions in the United State, Napa Valley is easily the most well-known for both its award-winning wines and **postcard-worthy** landscapes.*

Текстам статей закономерно свойственно обширное использование географических терминов. Рассмотрим их употребление на следующих примерах: *San Agustín is a small **mountain town** that's home to hundreds of pre-Columbian **statues and burial mounds**; Any trip to Norway should include a visit to ... **Norway fjords and fjord cruises**; Providencia is actually the center of a **huge***

*UNESCO Biosphere Reserve ...it contains some of the world's greatest **marine biodiversity**; But the main reason people visit it is to hike Los Nevados, a **mountain range** with majestic **snow-capped peaks**; Именно так называется красивейший **пещерный комплекс**...; Есть здесь и **баунти-острова** с пляжами, гигантские **пещеры**, китайская **архитектура**, и даже **гора-четырёхтысячник**; На **шиханах** построены **горнолыжные спуски**, **санатории** и **базы отдыха**.*

Анализ синтаксической структуры предложений в листиклах также позволил выявить некоторые характерные особенности. Так, в англоязычных заголовках в основном преобладают простые повествовательные предложения, распространенные определениями, выраженными в основном прилагательными, а также глаголами *travel, visit, see*: *50 Best Places to Travel in 2023; 10 Best Places to Travel on a Budget in 2025; 7 Must-see places to visit in Israel; The 23 Best Places to Travel in 2024; 16 Unmissable Places to Visit in Japan in 2025*. В русскоязычных заголовках, помимо простых повествовательных предложений, довольно частотны сложноподчиненные предложения, в которых зависимое предложение вводится словами *который, которые*: *Наши рекомендации: 15 отличных минских ресторанов, **которым** мы отдали свои сердца; 5 европейских городов, поездка в **которые** обойдётся меньше 20 тысяч рублей; 30 мест в России, **которые** обязательно стоит увидеть своими глазами; Глэмпинг в России: 20 мест, где можно почувствовать единение с природой; 10 мест планеты, **которые** должен посетить каждый!*

Во введении и основной части листикла наблюдается использование обращений к читателю, выраженных риторическими вопросами, что служит повышению вовлеченности читателя, помогает установить более личный контакт и расположить к себе аудиторию: *Are you on a whistle-stop tour of the capital or just **looking** for the **Berlin Top 10**? Are you **wondering where to go in Morocco** on your upcoming vacation?*

Помимо обращений к читателю в изучаемых статьях авторы также используют призывы к действию, выраженные побудительными предложениями с глаголами в повелительном наклонении. Подобный прием побуждает читателя не только пассивно воспринимать информацию, но и активно вовлекаться в прочтение и даже предпринимать конкретные действия. Стоит отметить, что при таких обращениях наблюдается частотное употребление глаголов в повелительном наклонении, таких как *discover, read on, take a look, explore, discover, miss, head to, читайте, арендуйте, отправляйтесь*: *Use this list to start planning your European getaway, and **cast** your vote below to have a say in next year's list; **Read more** about the places you need to see in Tallinn Old Town; Sample delicious sauvignon blanc in the Marlborough region or **explore** New Zealand's most accessible glaciers; When hunger strikes, **head to** the Mercato Centrale Firenze to stock up on fresh meats and cheeses or **sit down** for truffle tagliatelle or gelato at a charming outdoor eatery; In town, **don't miss** the chance to take the funicular; Итак, **читайте** про самые интересные страны для путешествий в 2025 году и путешествуйте!*

В русскоязычных статьях была зафиксирована синтаксическая конструкция, выраженная сложноподчиненным предложением с придаточным условия и главным побудительным предложением. Например, *Хотите X? Делайте Y* или *если X – делайте Y*. Рассмотрим употребление этих конструкций на примерах: *Если хочется экзотики, отправляйтесь на остров Борнео – ручаюсь, многие ваши друзья даже не знают, где это! Хотите совершите нетривиальное путешествие по Греции? Тогда следуйте нашему мари-руту: прилетайте в Салоники, арендуйте авто...*

Другой языковой особенностью является описание привлекательности и красоты объектов, рассматриваемых в основной части. Используются эмоциональные и образные описания и эпитеты, апеллирующие к чувствам читателя. В подобных вводных частях наблюдается обширное применение синтаксических конструкций с использованием сочинительных союзов с различными усилительными частицами: *Argentina: the land of wine, art and magnificent landscapes; From Patagonia to Buenos Aires, this beautiful country reaching to the tip of South America is one of the most inspiring in the world; В Беларуси много музеев: от монументальных с академическим подходом, как Национальный художественный музей в Минске, до уникальных, как столичный музей «Страна мини», в котором все ключевые белорусские достопримечательности можно увидеть за 2 часа в миниатюрах, а также камерные Музей Максима Богдановича или Музей Янки Купалы; Not only was it recognized by the Mexican government as a Magic Town in 2006, but it's also known for its ethereal jewel-toned lagoon...; Гора Филина находится в городе Лахденпохья и интересна не только живописными гранитными склонами, но и своим историческим значением; Гора Хуанишань – вечный источник вдохновения для поэтов и художников не только Китая, но и всего мира.*

Таким образом, основным жанрообразующим компонентом листикла является композиционная организация текста в виде нумерованного или маркированного списка. Пункты списка могут отмечаться маркерами вместо цифр, чтобы сделать акцент на том, что между объектами отсутствует различие по значимости. В сфере отдыха и туризма ранжирование в виде списка позволяет не преуменьшить значимость рассматриваемых объектов. Для жанра листикл также характерно обширное использование мультимедиа, таких как изображения и видеозаписи, и гиперссылок, с помощью которых читатель может получить дополнительную информацию. Особенностью листикла как жанра интернет-дискурса также является включение разговорных выражений. Активное использование разговорной лексики помогает сделать тексты понятными каждому пользователю, настраивает на дружеский лад, снижает официальность тона. Такой стиль повествования повышает вовлеченность читателя и создает ощущение диалога.

Подводя итог, можно сказать, что выбор языковых средств, используемых в листиклах, во многом связан с их тематической направленностью. Англоязычные и русскоязычные листиклы имеют больше сходств, чем

различий. Скорее всего, выявленные сходства связаны с общими тенденциями в SEO-оптимизации, интернет-журналистике и маркетинге, поскольку повышение читаемости текстов и вовлеченности аудитории, удобство в восприятии — это ключевые задачи. Стандарты, задаваемые глобальной интернет-средой, способствуют унификации форматов контента, что в значительной степени влияет на структуру и стиль статей листиклов, независимо от языка написания.

ЛИТЕРАТУРА

1. Белова К. А. Интернет-жанр: терминологическая вариативность и количественная неоднозначность // Иностранные языки: инновации, перспективы исследования и преподавания: материалы IV Междунар. науч.-практ. конф., посвящ. 100-летию Белорус. гос. ун-та, Респ. Беларусь, Минск, 26–27 марта 2021 г. / Белорус. гос. ун-т ; редкол.: Е. А. Пригодич (гл. ред.) [и др.]. – Минск : БГУ, 2021. – С. 42–47. URL: <https://elib.bsu.by/handle/123456789/262160> (дата обращения: 10.07.2025).
2. Деменьтев В. В. ‘n самых...’: интернет-рейтинг как речевой жанр // Жанры речи. 2021. № 3. С. 226–244.
3. Фролова В. В. Речевой жанр списка, его типы и жанрообразующие характеристики // Международный научно-исследовательский журнал. 2020. № 10-2 (100). С. 160–165. DOI: <https://doi.org/10.23670/IRJ.2020.100.10.067>.
4. Галичкина Е. Н. Интернет-дискурс: Основные направления изучения и тенденции развития // Известия ВГПУ. 2021. № 7. С. 94–101.
5. Ульянова М. А. Интернет-дискурс как жанр электронной коммуникации // Наука и современность. 2011. № 11. С. 349–354.

УДК 81-13 + 030 + 908

Солопова Ольга Александровна

доктор филологических наук,
профессор кафедры лингвистики
и перевода, старший научный сотруд-
ник управления научной и иннова-
ционной деятельности
Южно-Уральский государственный
университет (национальный исследо-
вательский университет)
г. Челябинск, Россия

Olga Solopova

PhD (Habil.) in Philology,
Professor at the Department
of Linguistics and Translation, Senior
Researcher at Research and Innovation
Services Laboratory
South Ural State University
(national research university)
Chelyabinsk, Russia
o-solopova@bk.ru

МАКРОСТРУКТУРА ПОЛИЯЗЫЧНОГО СЛОВАРЯ «БРИКС: ДИСКУРСИВНЫЕ МЕТАФОРЫ»

Структура словаря представляет единство формальных и содержательных принципов организации материала, направленных на реализацию конкретных лексикографических задач. Полиязычный словарь разрабатывается на материале пяти языков: русского, португальского, хинди, китайского, английского. Концепция словаря «БРИКС: дискурсивные метафоры» отвечает одной из глобальных целей лингвистики – изучению метафорической репрезентации общественно значимых феноменов на фоне исторического и лингвокультурологического контекстов. Данная цель реализуется в ряде частных задач: определение принципов ограничения исследуемого материала, отбор источников, фиксация доминантных фреймов, в рамках которых репрезентируется объединение; сплошная выборка метафор, сферой-мишенью которых является БРИКС в целом и входящие в объединение государства, отбор метафор для словника, разработка структуры словарной статьи. При разработке макро- и микроструктуры словаря используются процедуры автоматизированной и ручной выборки, корпусный и автоматизированный анализ, дефиниционный анализ, методика обобщения словарных дефиниций, фреймовый анализ, метод метафорического моделирования, лингвокультурный и когнитивно-дискурсивный методы. Целевая аудитория словаря представлена специалистами по политической лингвистике, журналистами, политологами, переводчиками.

Ключевые слова: лексикография; полиязычный словарь; корпус; медиадискурс; БРИКС; метафора; макроструктура.

Исследование выполнено за счет гранта Российского научного фонда № 24-18-00049, <https://rscf.ru/project/24-18-00049/>.

THE MACROSTRUCTURE OF THE MULTILINGUAL DICTIONARY “BRICS: DISCOURSE METAPHORS”

The dictionary structure represents a system of formal and content-based organizational principles designed to accomplish specific lexicographic objectives. Developed across five languages – Russian, Portuguese, Hindi, Chinese, and English, “BRICS: Discourse Metaphors” dictionary embodies one of the fundamental goals of linguistics: investigating the metaphorical representation of significant social issues within their historical and cultural contexts. The

conceptual framework is operationalized through several interrelated research tasks. They include establishing criteria for delimiting the study corpus, selecting appropriate textual sources, identifying dominant conceptual frames that structure representations of the BRICS grouping, sampling metaphors targeting BRICS and its constituent nations, selecting metaphors for the entry list, and designing the dictionary entry structure. Developing both the macro- and the microstructure employs an integrated methodological approach incorporating automated and manual text analysis procedures, definitional analysis, adapted techniques for summarizing dictionary definitions, frame analysis, metaphorical modelling, cultural, cognitive and discourse methods. The target audience comprises specialists in political linguistics, journalists, political analysts, and professional translators.

Key words: lexicography; multilingual dictionary; corpus; media discourse; BRICS; metaphor; macrostructure.

The research is financially supported by Russian Science Foundation № 24-18-00049, <https://rscf.ru/project/24-18-00049/>.

Макроструктура полиязычного словаря «БРИКС: дискурсивные метафоры» отвечает цели и задачам науки о теории и практике составления словарей на современном этапе развития во многих ее ипостасях: антропоцентрическая лексикография, цифровая лексикография, корпусная лексикография, фигуративная лексикография, объединенные более глобальным вектором развития науки – дискурсивная лексикография [1]. Дискурсивная лексикография постулирует мысль о том, что словарь должен отражать реальное употребление языка и быть предназначенным для конкретного пользователя [2].

Создание дискурсивного словаря метафор требует системного подхода к отбору языкового материала. В основу составления словаря положены три взаимодополняющих принципа: тематический, хронологический и частотный. Тематический принцип ориентирован на выявление метафор, репрезентирующих образы БРИКС и стран-участниц объединения [3]. Особое значение при этом приобретает анализ национальных медиадискурсов Бразилии, России, Индии, Китая, Южной Африки, позволяющий зафиксировать метафоры, отражающие процессы самоидентификации стран-участниц и конструирования образа БРИКС как особого геополитического проекта. Хронологический принцип предполагает ограничение материала периодом с 2006 г. по настоящее время (для южноафриканского дискурса нижняя граница хронологических рамок – 2011 г. – год официального включения Южно-Африканской Республики в объединение) [4]. Частотный принцип дает возможность зафиксировать релевантные концептуальные фреймы, сквозь призму которых осмысливается деятельность БРИКС [5].

Эмпирическую базу составили тексты электронных СМИ на пяти языках: португальском (Бразилия), русском (Россия), хинди (Индия), китайском (Китай) и английском (Южная Африка). Формирование корпусов осуществлялось с использованием автоматизированной выборки по ключевому слову *БРИКС* (на национальных языках). Объем каждого иллюстративного корпуса

составил 1 000 текстов, отсортированных по релевантности. После предварительной обработки, специфичной для каждого рабочего языка, применяются автоматизированные методы анализа текста с использованием специализированного обеспечения [6]: корпусные менеджеры НКРЯ (Национальный корпус русского языка) [7] (русский язык) и NOW (News on the Web) [8] (английский язык), автоматизированные программы: *Voyant Tools* [9] (португальский язык), *Sketch Engine* [10] (хинди), *AntConc* [11] (китайский язык).

Использование методов корпусной, компьютерной лингвистики и фреймового анализа позволяет выявить ключевые фреймы, в рамках которых репрезентируется БРИКС в медиадискурсе пяти государств, и установить степень их универсальности или специфичности [12]. Тем не менее, особенности разнотипных систем письменности рабочих языков (алфавитная, иероглифика, деванагари) и концептуальных единиц анализа (фреймов и метафор) требуют дальнейшей ручной верификации данных.

На последующем этапе осуществляется ручная обработка материала с применением метода метафорического моделирования, включающая идентификацию метафорических единиц, их фреймо-слотовый анализ, создание систем метафорических моделей, формирующих концептуальный «каркас» образов БРИКС и стран-участниц объединения в каждом национальном дискурсе, при сопоставительном анализе – выявление универсальных метафор, характерных для пяти дискурсов, и частотных национально-специфических единиц, типичных для нескольких [13; 14; 15; 16]. Соответственно, словарь состоит из двух частей, организованных по единым структурным принципам.

Все единицы, именуемые сферы-источники метафорических моделей, фреймов и слотов, являются существительными (например, *АРХИТЕКТУРА*, *БАРЬЕР*, *БИТВА*, *БЛОК (строительный)*, *БОЛЬ*, *БОРЬБА*, *БРАТ*, *БРАТСТВО*, *ВЕС*, *ВОЗДУХ*, *ВОРОТА*, *ДОБЫЧА*, *ДОМ*, *ДВЕРЬ*, *ДВИЖЕНИЕ*, *ДВИГАТЕЛЬ*, *ДРАЙВЕР*, *ДОРОГА*, *ДОРОЖНАЯ КАРТА*, *ДРУГ*, *ДРУЖБА*, *ГЛАВА (книги)*, *ГОРА*, *ЖИВОТНОЕ*, *ЗВЕЗДА*, *ЗВЕРЬ*, *ЗЕМЛЕТРЯСЕНИЕ* и др.).

Словарная статья содержит заголовочную единицу, зону семантизации, этимологические сведения, лингвокультурологический комментарий, иллюстративные примеры, дискурсивные смыслы, перекрестные ссылки. Все зоны приводятся для каждого из исследуемых языков и национальных дискурсов: в левой части словарной статьи – заголовочная единица (на одном из национальных языков БРИКС), в правой – остальные зоны в приведенной последовательности; зона семантизации является общей для всех языков. При наполнении словарной статьи используются дефиниционный анализ, методика обобщения словарных дефиниций, лингвокультурный и когнитивно-дискурсивный методы [17].

Словарная статья имеет четкую структуру: зоны визуально дифференцируются за счет особого шрифтового оформления. Проиллюстрируем наполнение словарных зон на примере заголовочной единицы *ПУТЬ*, универсальной для пяти дискурсов [18].

Заголовочная единица, представляющая сферу-источник метафоры, выделяется прописными буквами и полужирным шрифтом, при этом ее варианты на пяти языках разделяются косой чертой, например:

ПУТЬ / CAMINHO / मार्ग / 道路 / PATH(WAY)

Основное содержание набирается строчными буквами обычного шрифта. Зона семантизации включает нумерованные значения заголовочной единицы, общие для пяти языков, например:

1) Полоса земли, по которой осуществляется целенаправленное передвижение, сообщение. 2) Путешествие. 3) Направление движения.

Этимологическая справка приводится в том же шрифтовом оформлении, но без нумерации, например:

«**ПУТЬ** (рус.): пѣть (pѣть) ‘путь, дорога’ (ст.-слав.), pánthās ‘тропа, дорога, путь’ (др.-инд);

CAMINHO (порт.): camīo ‘путь, дорога’ (ст.-порт.), camminus ‘путь, дорога’ (позднелат.);

मार्ग (хинди): मार्ग (maarg) ‘путь, дорога, маршрут, направление, способ’ (санскр.), márgas ‘путь, след’ (праиндоиран.);

道路 (кит.): dàolù ‘дорога, путь’; 道 ‘путь; дорога; тракт’; 路 ‘дорога; путь, метод’ (традиц. кит.), 道 – 辵 ключевая графема ‘идти’, 路 – 足 ключевая графема ‘нога’;

PATH(WAY) (англ.): path + way (этимологическая тавтология), раѳ, раѳ ‘узкий проход, путь по земле, тропа, протоптанная человеком или животным’ (др. англ.); wei ‘проторенная дорога’ (ср. англ.), weg ‘тропа, путь, направление пути, расстояние, которое нужно преодолеть’ (др. англ.)».

В зоне лингвокультурологического комментария сохраняется строчный регистр и обычный шрифт, наименование лингвокультуры (например, *русская, бразильская* и др.) выделяется курсивом, например:

«В *русской* культуре путь – не только физическое движение к определенной цели, но и метафора развития, роста, поиска смысла жизни. Метафора «путь» в русскоязычной культуре – это универсальный, многослойный символ, отражающий жизненный процесс, нравственный выбор, испытания и стремление к цели. Она формирует особый тип мировосприятия, в котором движение, поиск и преодоление становятся центральными ценностями, а сама жизнь осмысливается как путь, наполненный смыслами и испытаниями. Носители русского языка связывают путь с движением к цели, поиском, испытаниями и развитием.

В *китайской* культуре в даосской традиции путь осмысливается как фундаментальный онтологический принцип, первооснова бытия и естественный порядок мироздания (Лао-цзы, «Дао Дэ Цзин»); в конфуцианстве путь обозначает нравственный императив следования добродетели («Лунь Юй»).

路 (lù) актуализирует репрезентирует как физические коммуникации (丝绸之路 «Шёлковый путь»), так и социальные траектории (人生路 «жизненный путь»). Графическая структура иероглифов кодирует идею поступательного развития, что находит отражение в современном политическом и массмедийном дискурсах (改革开放之路 «путь реформ и открытости»).

Иллюстративные примеры в словарной статье даются курсивом, при этом метафорическая единица выделяется полужирным шрифтом. Иллюстративные контексты сопровождаются указанием издания, даты выпуска, переводом на русский язык, например:

«БРИКС – **путь** к новым горизонтам стратегического партнерства (БКС-Экспресс, 04.09.2017).

*BRICS está pavimentando o **caminho** para a instalação da nova ordem global* (Central dos Trabalhadores Trabalhadoras do Brasil: СТВ, 24.10.2024). «БРИКС прокладывает **путь** к установлению нового глобального порядка».

ब्रिक्स: भारत का मध्यम मार्ग (Agniban, 27.06.2022). «БРИКС: срединный **путь** Индии».

金砖国家为繁荣铺平道路 (人民网, 01.07.2022). «Страны БРИКС прокладывают **путь** к процветанию».

*Such a commitment would enable a **pathway** towards building a stronger future of BRICS* (ЮАР, News 24, 22.07.2018). «Подобное обязательство проложило бы **путь** к построению более прочного будущего БРИКС».

В зоне дискурсивного комментария, набираемой строчными буквами и полужирным шрифтом, акцентируются смыслы, которые метафора транслирует в медиадискурсах стран БРИКС, например, для метафоры ПУТЬ эта зона представлена следующими дискурсивными смыслами: «развитие, поступательность, целенаправленность, преодоление, альтернатива».

Словарная статья для метафоры на одном из рабочих языков, включающая все вышеперечисленные зоны, построена следующим образом:

«**मार्ग** (хинди): मार्ग (maarg) ‘путь, дорога, маршрут, направление, способ’ (санскр.), márgas ‘путь, след’ (праиндоиран.). В индийской культуре «maarg» используется для обозначения места в пространстве, по которому осуществляется передвижение; возможности достижения некой цели; духовного развития, в частности путешествия к постижению Бога. Например, в буддизме и индуизме, религиях, которые зародились на Индийском субконтиненте, под «maarg» понимается путь к достижению мокши, т. е. спасения от круговорота перерождений. В важнейшем тексте индуистской философии Бхагавадгите говорится о том, что основным способом достижения мокши является следование по тройственному пути (тримарга), основными элементами которого являются: джняна-марга или «путь знания» (изучение философских текстов и размышлений), карма-марга, «путь действия» (надлежащее исполнение рели-

гиозных и этических обязанностей), и бхакти-марга, «путь преданности и самоотдачи Богу». *ब्रिक्स: भारत का मध्यम मार्ग* (Agniban, 27.06.2022). **БРИКС: срединный путь Индии».**

Для разрешения фундаментального противоречия между алфавитной организацией материала в словаре, обеспечивающей оперативный поиск нужной единицы, соблюдение стандартов лексикографической практики и удобство работы со словарем как справочным изданием, и системной природой метафорических моделей, каждая словарная статья завершается перекрестными ссылками на метафоры, концептуально связанными с рассмотренной. Например, для заголовочной единицы *ПУТЬ* в конце словарной статьи приведен блок перекрестных ссылок, набранных прописными буквами обычного шрифта с двойным пробелом между единицами:

«БАРЬЕР, ДВИЖЕНИЕ, ДОРОГА, ДОРОЖНАЯ КАРТА, МАГИСТРАЛЬ, ПЕРВОПРОХОДЕЦ, ПЕРЕКРЕСТОК, ПЕРЕХОД, ПОПУТЧИК, ПУТЕШЕСТВИЕ, ПУНКТ НАЗНАЧЕНИЯ, СКОРОСТЬ, ТРАЕКТОРИЯ, ШАГ».

Лексикографическое описание заголовочных единиц для каждого национального дискурса, зоны дискурсивных смыслов и перекрестных ссылок отделяются друг от друга пустой строкой, что существенно улучшает идентификацию различных типов информации и навигацию по словарной статье.

Характеристика макроструктуры словаря была бы неполной без упоминания о системе помет и сокращений, например:

- англ. — английский
- г. — год
- гл. — глагол
- греч. — греческий
- досл. — дословно
- др. англ. — древнеанглийский
- з. сакс. — западно-саксонский
- кл. лат. — классическая латынь
- лат. — латинский

Таким образом, словарь «БРИКС: дискурсивные метафоры» решает несколько задач теоретического и прикладного характера. Во-первых, такой словарь позволяет продемонстрировать когнитивно-дискурсивный потенциал метафорической репрезентации действительности, когда репертуар образных средств определяется как типологическими характеристиками языков, так и культурными особенностями категоризации и концептуализации мира. Во-вторых, такой лексикографический продукт дает возможность зафиксировать не нормативное, а реальное употребление метафор в медиадискурсе, который отражает социально-политические изменения не постфактум, а в момент функционирования в рамках того или иного лингвокультурного сообщества. В-третьих, данный словарь наглядно показывает, что лингвистика в настоя-

щее время расширяет репертуар исследовательских методов, обращаясь к корпусным инструментариям и автоматизированным программам с целью более точного представления языковой ситуации в ее экстралингвистическом преломлении.

ЛИТЕРАТУРА

1. Кошкарлова Н. Н., Солопова О. А. Проект словаря метафор «Образ России в медиадискурсе стран БРИКС» // Лексикография цифровой эпохи: Сборник материалов II Международного симпозиума, Москва-Томск, 20–22 ноября 2024 года. Томск : Национальный исследовательский Томский государственный университет, 2024. С. 93–96.
2. Ловчикова А. А., Солопова О. А. Дискурсивная лексикография: перспективы изучения политической метафоры // Вестник Южно-Уральского государственного университета. Сер. Лингвистика. 2024. Т. 21. № 2. С. 28–35.
3. Керимов А. А. БРИКС: состояние, проблемы и перспективы // Известия Иркутского государственного университета. Сер. Политология. Религиоведение. 2024. Т. 50. С. 52–61. <https://doi.org/10.26516/2073-3380.2024.50.52>.
4. Сибиряков И. В. Рождение БРИК: первые программные заявления новой международной организации // Вестник Южно-Уральского государственного университета. Сер. Социально-гуманитарные науки. 2024. Т. 24. № 4. С. 43–48.
5. Солопова О. А. Образ БРИКС в медиадискурсе ЮАР: диагностические фреймы // Язык в эпоху цифровых трансформаций и развития искусственного интеллекта: сб. науч. ст. по итогам междунар. науч. конф., Минск, 23–24 окт. 2024 г. / Редкол.: Н. Е. Лаптева (отв. ред.) [и др.]. Минск : МГЛУ, 2024. С. 610–617.
6. Парулина И. Ю., Солопова О. А., Кошкарлова Н. Н. Медиадискурсы стран БРИКС: корпусные и компьютерные технологии в исследовании языка // Право в эпоху искусственного интеллекта: перспективные вызовы и современные задачи: сборник научных статей по материалам Международного научно-практического форума, Тюмень, 17–19 октября 2024 года. Тюмень : ТюмГУ-Press, 2024. С. 62–67.
7. Национальный корпус русского языка (НКРЯ). URL: <https://ruscorpora.ru> (дата обращения: 08.04.2024).
8. News on the Web (NOW). URL: <https://www.english-corpora.org/now/> (дата обращения: 21.08.2024).
9. Voyant Tools. URL: <https://voyant-tools.org> (дата обращения: 26.06.2024).
10. Sketch Engine. URL: <https://www.sketchengine.eu> (дата обращения: 09.08.2024).
11. AntConc. URL: <https://www.laurenceanthony.net/software/antconc/> (дата обращения: 23.07.2024).
12. Солопова О. А., Кошкарлова Н. Н. Доброе братство лучше богатства: семейная метафора в дискурсах стран БРИКС // Южно-Российский журнал социальных наук. 2024. Т. 25. № 4. С. 39–54. <https://doi.org/10.31429/26190567-25-4-39-54>.

13. Голубцова Е. В. Языковые средства манипуляции западных СМИ при формировании образа России (на материале BBC Hindi и Deutsche Welle Hindi) // Вопросы журналистики, педагогики, языкознания. 2025. № 44 (1). С. 138–149. <https://doi.org/10.52575/2712-7451-2025-44-1-138-149>.
14. Информационная война: формы ведения и методы лингвистического анализа: коллективная монография / Под ред. О. И. Калинина, О. А. Солоповой. Москва : ФЛИНТА, 2025. 344 с.
15. Калинин О. И. Образ России в дискурсе социальных сетей Китая: лингво-прагматический анализ репрезентации российско-украинского конфликта // Вестник Северного (Арктического) федерального университета. Сер. Гуманитарные и социальные науки. 2024. Т. 24. № 6. С. 83–91. <https://doi.org/10.37482/2687-1505-V392>.
16. Солопова О. А., Кашяп Г., Нагаев Н. С. Метафорический образ России в медиадискурсе Индии // Политическая лингвистика. 2024. № 4 (106). С. 89–97.
17. Солопова О. А., Кошкарлова Н. Н. Микроструктура полиязычного словаря «БРИКС: дискурсивные метафоры» // Вопросы лексикографии. 2025. № 36. С. 26–44. doi: 10.17223/22274200/36/2
18. Кошкарлова Н. Н., Солопова О. А., Чудинов А. П. На пути к многополярному миру: метафоры в декларациях саммитов БРИКС. Полис. Политические исследования. 2025. № 3. С. 113–124. <https://doi.org/10.17976/jpps/2025.03.08>. EDN: KEEQQB.

УДК 004.89

Тарасенко Екатерина Анатольевна
студентка

Академия управления при Президенте
Республики Беларусь
г. Минск, Беларусь

Ekaterina Tarasenko

Student
Academy of Public Administration
under the aegis of the President
of the Republic of Belarus
Minsk, Belarus
katytara2004@gmail.com

Петренко Елизавета Васильевна
студентка

Академия управления при Президенте
Республики Беларусь
г. Минск, Беларусь

Elizaveta Petrenko

Student
Academy of Public Administration
under the aegis of the President
of the Republic of Belarus
Minsk, Belarus
lizapetrenko13@gmail.com

Кудрейко Татьяна Владимировна
старший преподаватель

Академия управления при Президенте
Республики Беларусь
г. Минск, Беларусь

Tatyana Kudreyko

Senior lecturer
Academy of Public Administration
under the aegis of the President
of the Republic of Belarus
Minsk, Belarus
trusakevich@yandex.by

ПРОБЛЕМА АВТОРСТВА И ПРАВООБЛАДАНИЯ В ИИ-СГЕНЕРИРОВАННЫХ ТЕКСТАХ

С развитием генеративных моделей искусственного интеллекта (ИИ), способных создавать связные и осмысленные тексты, остро встал вопрос об авторстве и праве на результаты такой генерации. Кто является автором текста, созданного ИИ? Если модель обучена на чужих работах, чьи идеи она транслирует? И каково место этих текстов в правовой и лингвистической системах? В данной статье анализируются ключевые подходы к понятию авторства в лингвистике и философии, рассматриваются международные правовые прецеденты и различия в регулировании в США, Великобритании, Европейском союзе, Китае, Японии и Беларуси. Отдельное внимание уделено проблеме использования защищенных авторским правом материалов для обучения ИИ и возможным механизмам правового урегулирования, включая лицензирование данных и обеспечение прозрачности источников. Кроме того, акцентируется роль человека в процессе генерации: обсуждается степень вовлеченности пользователя и разработчика, а также границы их творческого вклада. Подчеркивается необходимость разработки новых юридических и этических норм, которые позволят сбалансировать интересы пользователей, разработчиков и правообладателей, сохраняя при этом роль человека как основного носителя ответственности и творческого замысла.

Ключевые слова: искусственный интеллект; авторство; авторское право; нейросеть; ИИ-контент.

THE PROBLEM OF AUTHORSHIP AND COPYRIGHT IN AI-GENERATED TEXTS

The rapid development of generative artificial intelligence (AI) models, capable of producing coherent and meaningful texts, has sharply raised questions about authorship and copyright over the results of such generation. Who is the author of a text created by AI? If the model is trained on others' works, whose ideas does it convey? What is the place of these texts within the legal and linguistic system? This article analyzes key approaches to the concept of authorship in linguistics and philosophy, examines international legal precedents and considers differences in regulation in the United States, the United Kingdom, the European Union, China, Japan and Belarus. Special attention is given to the use of copyrighted materials for AI training and to possible mechanisms of legal regulation including data licensing and the transparency of training sources. In addition, the role of human involvement in the generation process is discussed, with particular focus on the boundaries of creative contribution by users and developers. The article also addresses the boundaries of human involvement in the generation process and the criteria for recognizing the contribution of a user or developer as creative. It emphasizes the need to develop new legal and ethical norms that balance the interests of users, developers and copyright holders while preserving the role of the human being as the primary bearer of responsibility and creative intent.

Key words: artificial intelligence; authorship; copyright; neural network; AI-generated content.

The problem of authorship remains one of the most crucial and ambiguous in contemporary humanities. It is not limited to the idea of a specific person who created a text; it can be understood as a source of intention, a bearer of a specific style and text structure, or a cultural mechanism that shapes the content and form of a work. The category of authorship is considered in various interpretations. For instance, V. V. Vinogradov introduced the term "image of the author" as a reflection of personality in the text's style. M. M. Bakhtin viewed the author as a creative force expressing a worldview and artistic vision of the world; not merely a creator, but an active participant in the "event of being". Rodnyanskaya highlights the figure of the "author-narrator," while Kozhinov speaks of the "author's voice" as an expression of ideological and emotional layers of utterance [1]. Thus, an author is not always a specific individual, but invariably a source of intention and the semantic dominant of the text.

It is precisely at this point that difficulties arise in applying classical notions of authorship to texts created by artificial intelligence. AI lacks subjectivity, worldview, or intention – properties traditionally considered essential attributes of an author. Classical linguistics views the author as a discursive function and organizer of meaning; however, AI possesses neither intention nor self-awareness. It does not create; it generates. Its texts do not stem from communicative necessity but are statistically conditioned probabilistic constructions based on training data.

This casts doubt on the applicability of the traditional term "author" to machine generation. In such a case, who is responsible for the content of the text:

the algorithm developer, the user who initiated the generation, or none of them? In conditions of blurred concepts and the absence of creative subjectivity in artificial intelligence, there arises a need to rethink the very concept of authorship in relation to new technological conditions. AI does not make conscious decisions, is not responsible for the content of the generated text, and does not understand the meaning of the generated result. However, a user who merely inputs a query without significant refinement of the obtained material cannot always be considered a full-fledged author in the classical sense. The developers, despite creating the architecture of the model and training it on large amounts of data, do not directly participate in the generation process of a specific text. All of this creates a philosophical and legal "grey area" where none of the participants fully fit the traditional definitions of an author.

A striking example of these difficulties was the case of the artist Jason Allen, who in 2022 won a fine art competition in Colorado with his image "Théâtre d'Opéra Spatial," created using the Midjourney neural network [2]. Allen spent over a hundred hours formulating and refining more than 600 prompts, polishing the image in a graphics editor, and scaling it using the Gigapixel AI tool. Despite his large-scale creative contribution, the U.S. Copyright Office denied his copyright registration in 2025, citing the lack of "human authorship" – most of the visual content, in their opinion, was created by AI. Allen filed a lawsuit in federal court, arguing that his conception, guidance, selection and adjustment of prompts, as well as subsequent image processing, constituted a form of creative control deserving legal protection. This case became one of the first high-profile precedents vividly demonstrating the gap between the capabilities of modern AI systems and the limits of existing copyright legislation. It caused a wide public response and stimulated discussions about the need to update legal norms, as well as the degree of human participation in the creative process when using generative technologies.

As a result, artificial intelligence should be considered not as an author, but as a tool capable of simulating authorship. However, in situations where AI creates text or an image without significant human involvement, the question arises: is it possible to attribute authorship to the developers of the model or its users? Until legal practice has developed a unified approach, the concept of authorship remains open to interpretation, especially in an era when the boundaries between an instrument and a co-author are rapidly blurring.

An additional complexity is introduced by the way AI models themselves are formed, in particular, the issue of the legality of using training data. Modern language models are trained on large amounts of texts, often obtained without the direct consent of copyright holders. This has led to prominent legal disputes. For example, in 2023, Getty Images sued Stability AI, claiming that it used millions of copyrighted images to train its models [3]. The case highlights an important legal

dilemma: is AI training considered "fair use" or copyright infringement? Although texts and images are regulated slightly differently, the precedent is relevant for language models as well: the use of copyrighted books, articles, and media may be considered illegitimate copying if there is no transparent licensing mechanism. In this regard, various proposals for regulating and legalizing the use of data for training artificial intelligence are on the agenda.

One of the key solutions is to implement data licensing systems similar to those used in the music industry: authors receive payments for each instance of their works being used in the model training process. Such a model makes it possible to balance the interests of copyright holders and developers of artificial intelligence, ensuring fair compensation and stimulating further creativity. An example of this approach is the Spawning AI platform [4], which allows authors to opt-out of having their content used, increasing transparency and respect for creators' rights.

Transparency of models is also becoming a mandatory element of regulation. The legislation of the European Union, which is scheduled to be adopted in 2025, requires disclosure of the sources of training data, which will allow copyright holders and users to understand what materials the models were trained on [5]. Additionally, the introduction of content filtering by style is being discussed, which would allow blocking the generation of images or texts in the manner of specific authors without their consent, thereby protecting the uniqueness and rights of creators. Such measures reflect a desire for clearer regulation, but uniform standards do not yet exist.

The approach to copyright for AI-generated content varies significantly from country to country. For example, in the United States, in the *Thaler vs. USPTO* case (2022), the court ruled that only a person can be the author of a work [6]. Content created by AI without significant human creative input is not subject to copyright protection. This means that such content is in the public domain, and no one can claim exclusive rights. In the UK, there is also an approach according to which only a person can be the author of a work. If the content is created using a computer without human involvement in the creative process, the copyright is assigned to the one who initiated the creation, for example, the developer or the owner of the program. At the same time, the law refers to the "person who took the necessary measures" to generate, recognizing him as the copyright holder. In the European Union, the issue of authorship in the context of AI remains the subject of active discussion. The 2020 Resolution of the European Parliament highlights the difference between works created with significant human involvement and those generated by AI autonomously [7]. In the first case, the rights to the results are recognized for the person as a full-fledged author. In the second case, when AI creates content without significant human creative input, it is proposed to consider the possibility of introducing a new category of intellectual property. This will

ensure the legal protection of such facilities and determine who may own the relevant rights. At the same time, in some EU countries, the issue of mandatory transparency of the sources from which AI systems are trained is being raised, which further complicates legal regulation.

China holds one of the most progressive positions in Asia regarding the protection of neural network-generated works. There, copyright protection for AI-created content is permissible if human creative contribution can be proven. For example, in the case of *Tencent v. Yingxun*, the Chinese court ruled that an article created with the help of artificial intelligence has originality and is subject to copyright protection [8]. Japan, in contrast, maintains a more restrained position. According to Japan's Copyright Law, a work must express "human thoughts and emotions". Nevertheless, in January 2024, the Japanese ministries issued guidelines that allow copyright protection of works created with the participation of AI, provided significant human contribution [9]. Criteria considered include: the level of detail in prompts, the number of generation attempts, and participation in selecting or refining the result. In such cases, the author is recognized as the person who has made a decisive creative contribution.

As for Belarus, according to the position of local legal experts, artificial intelligence cannot be considered as an author within the framework of current legislation. Nevertheless, if a work is created as a result of active human interaction with AI, copyright may be assigned to the user as the primary subject of creativity. At the same time, AI developers can claim a certain share of the rights, especially if their contribution to the result turned out to be significant – such cases are regulated by licensing agreements. As lawyers from the National Center for Intellectual Property emphasize, it is a person who formulates ideas and directs AI, while the model itself acts only as a tool that analyzes data and produces a result [10]. In situations of close cooperation between the user and the system, a flexible approach to the allocation of rights is possible, taking into account the contribution of both parties.

Thus, the problem of authorship and copyright ownership in the context of AI-generated texts reflects a deep crisis of traditional humanitarian and legal categories. Modern technologies challenge the established notion of the author as a subject of intention, creativity and responsibility. Artificial intelligence, in essence, is not a subject capable of creativity in the classical sense: it does not think, does not interpret, does not realize what is being created. Nevertheless, the results of its work are becoming increasingly indistinguishable from human-made works. Global legal practice demonstrates the lack of a unified approach to determining authorship in such cases: from the complete denial of copyrights to AI content to attempts to flexibly interpret human contributions to the generation process. This confirms the need to review existing legislation and develop new legal categories that take into account the specifics of human-machine interaction. Linguistic and

philosophical approaches also point to the need to separate the concepts of "author" and "tool" – while recognizing that AI can serve as a powerful mediator in realizing human intent.

In the context of the rapid development of technology, transparency of training data, ethics of using other people's works, as well as a fair distribution of rights among all participants in the creative process are of particular importance. AI should not be considered as an autonomous author, but as an intellectual tool, the value of which is revealed only in tandem with humans. And it is the human being – the user, the curator, the developer – who remains the primary bearer of responsibility, intention and, ultimately, rights.

LITERATURE

1. Бахтин М. М. Эстетика словесного творчества. – 2-е изд. – М.: Искусство, 1986. – 445 с.
2. Who Owns the Copyright for an AI-Generated Artwork? [Electronic resource]. URL: <https://hyperallergic.com/957357/who-owns-the-copyright-for-an-ai-generated-artwork/> (accessed: 07.06.2025).
3. Getty Images is suing the creators of AI art tool Stable Diffusion for scraping its content [Electronic resource]. URL: <https://www.theverge.com/2023/1/17/23558516/ai-art-copyright-stable-diffusion-getty-images-lawsuit> (accessed: 01.06.2025).
4. Data Governance for Generative AI [Electronic resource]. URL: <https://spawning.ai/> (accessed: 09.06.2025).
5. EU AI Act: New Disclosure Template Might Spark Copyright Disputes [Electronic resource]. URL: <https://www.gamingtechlaw.com/2025/01/eu-ai-act-new-disclosure-template-sparks-copyright-disputes/> (accessed: 09.06.2025).
6. US Supreme Court rejects computer scientist's lawsuit over AI-generated inventions [Electronic resource]. URL: <https://www.reuters.com/legal/us-supreme-court-rejects-computer-scientists-lawsuit-over-ai-generated-2023-04-24/> (accessed: 05.06.2025).
7. European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies [Electronic resource]. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.html (accessed: 05.06.2025).
8. Tencent v. Yingxun Tech [Electronic resource]. URL: <https://www.chinajusticeobserver.com/law/x/2019-yue-0305-min-chu-14010> (accessed: 05.06.2025).
9. AI Guidelines for Business [Electronic resource]. URL: https://www.soumu.go.jp/main_content/000916810.pdf (accessed: 05.06.2025).
10. Как распределяются права при создании произведений с помощью ИИ, рассказали в НЦИС [Электронный ресурс]. URL: <https://pravo.by/novosti/obshchestvenno-politicheskie-i-v-oblasti-prava/2025/april/88484/> (дата обращения: 07.06.2025).

УДК 811.134

Холова Зарина Дилшодовна

ассистент

РУДН им. Патриса Лумумбы

г. Москва, Россия

Zarina Kholova

Assistant

RUDN named after Patrice Lumumba

Moscow, Russia

Kholova_zd@pfur.ru

ПСИХОЛИНГВИСТИЧЕСКИЙ АНАЛИЗ ЛИНГВОКУЛЬТУРНЫХ ОСОБЕННОСТЕЙ КОРОТКИХ ВИДЕОРОЛИКОВ (ШОРТСОВ) В РУССКОЯЗЫЧНОМ И ИСПАНОЯЗЫЧНОМ ЦИФРОВОМ ДИСКУРСЕ

В данной статье представлен психолингвистический анализ языковых и культурных особенностей коротких видеороликов (Shorts) в русскоязычном и испаноязычном цифровом дискурсе. Исследование рассматривает специфические лексические, синтаксические и прагматические характеристики Shorts, а также их культурные и коммуникативные функции. Сравнительный анализ позволяет выявить сходства и различия в способах конструирования смысла, выражения эмоций и взаимодействия с аудиторией в данном цифровом формате. Результаты исследования способствуют более глубокому пониманию взаимосвязи языка, культуры и цифровой коммуникации, подчеркивая роль Shorts как инструмента межкультурного обмена и современного медиапотребления.

Ключевые слова: короткие видео; медиадискурс; мультимодальность; лингвокультурные особенности; кросс-культурная прагматика.

PSYCHOLINGUISTIC ANALYSIS OF LINGUISTIC AND CULTURAL FEATURES OF SHORT VIDEOS (SHORTS) IN RUSSIAN-LANGUAGE AND SPANISH-LANGUAGE DIGITAL DISCOURSE

This article presents a psycholinguistic analysis of the linguistic and cultural features of short videos (Shorts) in Russian-language and Spanish-language digital discourse. The study examines the specific lexical, syntactic, and pragmatic characteristics of Shorts, as well as their cultural and communicative functions. Through comparative analysis, the research identifies similarities and differences in the ways Russian and Spanish speakers construct meaning, express emotions, and engage with audiences in this digital format. The findings contribute to a deeper understanding of the interaction between language, culture, and digital communication, highlighting the role of Shorts as a tool for intercultural exchange and modern media consumption.

Key words: shorts; media discourse; multimodality; linguacultural features; cross-cultural pragmatics.

Парадигма новых медиа не только изменила медиаландшафт, но и повлияла на развитие межкультурной коммуникации, упростив процесс и сократив дистанцию между индивидами. Так, современные цифровые платформы, где представлены короткие видео, или шортсы (*shorts* – короткие видео),

(к примеру, TikTok и YouTube) представляют собой место для экспрессии культурных особенностей. Шортсы выполняют несколько функций, в том числе информационную, коммуникационную, рекреационную и культурную.

Так, короткие видео, развлекая, отражают лингвокультурные особенности разных наций, этносов и народов. Здесь уместно исследование лингвокультурных дифференциаций шортсов на русском и испанском языках с целью выявления особенностей в языковых средствах, невербальной коммуникации и культурных ценностях.

Важно обратиться в первую очередь к понятию лингвокультурологии. Данная дисциплина, которая находится на стыке науки о языке и науки о культуре, изучает взаимосвязь между этими феноменами. Лингвокультурология сосредоточена на исследовании «языковой или дискурсивной деятельности, рассматриваемой с ценностно-смысловой точки зрения» [2, с. 121].

В контексте шортсов лингвокультурологический фокус изучения держится на такие вербальные и невербальные средства и способы коммуникации, как лексические средства, интонация, жесты, тематика, музыкальное сопровождение, колористика и пр. Отмечается, что культурные различия влияют на восприятие и интерпретацию материала, что делает лингвокультурный анализ необходимым для понимания межкультурной коммуникации.

Короткие видео выступают как цифровые инструменты, которые полезны в разрезе передачи информации о культурных нормах и ценностях индивида и некоего общества [1]. Особенность шортсов состоит в том, что они успешно сочетают глобальные стили и тенденции (которые называют трендами) и местных традиций, и элементов менталитета. Короткие видео позволяют гражданам разных стран принимать участие в процессах мировой медиакультуры и обмена ценностями.

Стоит перейти к непосредственному анализу лингвокультурных различий, которые можно найти в коротких видео на русском и испанском языках. Итак, в рамках исследования были выбраны такие платформы, как TikTok [5] и YouTube Shorts [6]. Можно выделить некоторые особенности шортсов в русскоязычном сегменте. Контент на русском языке отличается использованием таких литературных приемов и средств, как сарказм, ирония, игра слов, благодаря чему пользователи могут создавать комичные, поучительные ситуации. Видео часто носят репрезентативное назначение, то есть они призваны показать реальные жизненные ситуации, которые понятны большинству или многим. Так, например, часто русскоязычные создатели шортсов используют типичные фразы для описания контекста. Популярным текстом является следующий: «*понимание: 0 % // вайб: 100 %*» [6]. Например, пользовательница @ddmchkk в YouTube Shorts использует данные лексические единицы под липсинг песни об абьюзивных отношениях и намекает на политическую ситуацию [6].

Также следует отметить такую лингвокультурную особенность русскоязычных шортсов политической тематики, как использование шуток. К примеру, в коротком видео @Rf_Leader на платформе YouTube Shorts приводится шутка Президента Российской Федерации Владимира Владимировича Путина, где он уточняет у блогера Влада А4, является ли «Влад Бумага» псевдонимом молодого человека. Получив положительный ответ, В. В. Путин делает отсылку к игре камень, ножницы, бумага: *«Отлично, бумага, как в известной игре, сильнее камня»* [6]. Так, шортс подчеркивает такую особенность российской культуры (в лице главы Российского государства), как игровой элемент, склонность шутить.

В связи с особой геополитической ситуацией, особенно часто в политической реальности коротких видео упоминаются фамилии политиков, президентов стран, выполняющих ключевую функцию в решение актуальных экономико-политических конфликтов. Так, фамилия Трамп приобретает особую популярность в контексте шортсов на русском языке.

Однако надо заметить, что обсуждение Президента США Дональда Трампа проходят не всегда с позитивной точки зрения. К примеру, в коротком видео @TNT_television в YouTube Shorts высмеивается санкционная политика США против Российской Федерации: в 2024 году бывший Президент Джо Байден ввел санкции относительно экспорта некоторых товаров в Россию: *«В последнем пакете я им запретил кроссовки Джордан. Я же говорю: я в тупике. У меня сейчас указы не как у президента, а как у, не знаю, ведущего «Модного приговора»*. Трамп, который входит в Овальный кабинет, был тем самым *«одним человечком, который заставит этих русских молить о пощаде и новеньких джорданах»* [6].

Комментируя данный шортс, можно обратить внимание на снижение официальной стилистики, которое, в свою очередь, придает политическому дискурсу ироничный, даже гротескный характер. Сленговая форма в данном контексте не только маркирует принадлежность к массовой культуре, но и способствует эмоциональной идентификации зрителя с комментатором – это типичный прием вовлечения в рамках интернет-коммуникации. Здесь благодаря использованию разговорной лексики («джорданы»), аллюзий («Модный приговор»), которые выступают в качестве культурных маркеров, устойчивых выражений («молить о пощаде»), уменьшительно-ласкательного суффикса -чек- («человечек») демонстрируется не только богатство русского языка, но и специфика русскоязычного восприятия и интерпретации международной повестки, в частности – образа американских лидеров и актуальной политической повестки [6]. Само короткое видео является типичным в русскоязычном интернет-пространстве, поскольку обыгрывается ироничный и критический подход, в том числе к политике. Подобная практика иллюстрирует то, как лингвокультурные особенности языка формируют восприятие глобальных событий в масс-медиа.

Говоря об особенностях русской культуры, стоит заметить, что эта культура склонна к высокой степени толерантности к неоднозначности, что проявляется в принятии парадоксальных, ироничных высказываний о власти [4]. Это объясняет популярность сатирических роликов, в которых мировые лидеры становятся объектами пародий.

Представляется возможным перейти к испаноязычным коротким видео. Для анализа используется контент политического характера, который имеет отношение к Боливарианской Республике Венесуэла. В коротком видео пользователя @Libre269 в TikTok приводится цитата Президента Венесуэлы Николаса Мадуро, где он довольно эмоционально заявляется о сложной геополитической ситуации: ... *Denunciando hasta el Consejo de Seguridad de la ONU que a Rusia pretenden rodearla, apuntar todas las armas de la OTAN sobre Rusia para en algún momento que se presenten las condiciones atacar y destruir a Rusia y que pretende el mundo* (перевод: «... Осуждая вплоть до Совета Безопасности ООН, что они намереваются окружить Россию, направить все оружие НАТО на Россию, чтобы в какой-то момент создать условия для нападения и уничтожения России, что и намеревается сделать мир») [5].

Данное короткое видео отражает такие культурные ценности Венесуэлы, как дружба с Россией и уважение к прошлому, где СССР поддерживал теплые отношения с Венесуэлой. Также следует отметить особую эмоциональность, которая является особенностью испаноязычных обществ, которые характеризуются высокой экспрессивностью и эмоциональностью коммуникации. Используя эмоционально насыщенные глаголы, такие как *rodearla*, *apuntar*, *atacar*, *destruir* («окружить ее», «прицелиться», «атаковать», «уничтожить»), Президент повышает градус драматичности ситуации, сопереживая партнеру [5]. Интонационно насыщенная речь является особенностью публичных выступлений на испанском языке и является элементом культуры и национальной риторической традиции [4]. Отдельно стоит заметить, что речь произносит мужчина, который, очевидно, находится в сильных чувствах отрицательного характера, а его страстность подразумевается в качестве проявления искренности и силы позиции, что соответствует особой роли маскулинности в венесуэльской культуре.

Также в испаноязычных шортах политической тематики можно часто встретить обсуждения политических мероприятий. Например, в коротком видео BRICS OFICIAL в TikTok идет речь о впечатлении Дмитрия Николаевича Чернышенко, заместителя Председателя Правительства РФ, о сотрудничестве России с Венесуэлой [5].

По поиску на платформе YouTube Shorts по ключевым словам *Venezuela politics* можно часто встретить короткие видео, которые посвящены проблеме депортации венесуэльцев из США, их возвращении домой, кризисе в государстве. Данные видео являются попыткой обсудить актуальные социальные и политические темы в понятном большинству формате, который пользуется

особой популярностью. Упомянутые тему отражают реальные боли граждан Венесуэлы и используются в качестве элементов эмоционального вовлечения: короткие видео на испанском языке сфокусированы на такой культурной особенности венесуэльского менталитета, как эмоциональность. Так, делая акцент на персональном опыте и истории личности, авторы таких шортсов делают из жертв социально-политических несправедливостей героев, которые много чувствуют и много говорят.

Можно выделить невербальные символы и визуальные метафоры, такие как плач, образы семьи, разлуки, границ. Формирование образа коллективной боли и борьбы происходит благодаря использованию личных местоимений, ключевых слов, таких как *deportación* («депортация»), *nuestro país* («наша страна») и др. Часто на фоне играет музыка на испанском языке, которая содержит национальный колорит, – так авторы шортсов отправляют конкретные послышки целевой аудитории: этот страдающий венесуэлец – один из нас, представитель нашей страдающей нации.

Так, короткие видео на русском языке часто опираются на иронию, сарказм, языковую игру, что отражает высокую толерантность к неопределенности в культуре. Политические шортсы также демонстрируют критическое восприятие международной повестки, а применение разговорной лексики, аллюзий и суффиксов подчеркивает приближение официального дискурса к массовой культуре. Данные средства позволяют создать эффект социального комментирования через юмор, характерный для русской медиакультуры.

В испаноязычных коротких видео отмечается эмоциональный и драматичный стиль, где используется интонационная экспрессивность и лексическая насыщенность глаголами действия. Высокая эмоциональность речи соответствует традициям публичной риторики и эмоционального выражения. Испаноязычные короткие видео часто касаются таких острых социально-политических тем, как миграция, кризис, социальные переживания. Короткие видео на испанском языке часто содержат такие средства, как плач, разлука, семья, музыка с национальным колоритом, которые выполняют инструментальную роль при формировании национальной венесуэльской идентичности страдающего народа, а сами короткие видео репрезентует коллективный катарсис

Таким образом, лингвокультурные различия в коротких видео на русском и испанском языках можно отследить благодаря анализу языковых особенностей, невербальных средств и тематики контента.

ЛИТЕРАТУРА

1. Анисимов, В. Е. Шортс, рилс и клип как субвидеа интернет-, кино- и медиадискурса / В. Е. Анисимов, Е. Д. Анисимова // Филологические науки. Вопросы теории и практики. – 2024. – Т. 17, № 7. – С. 2386–2398.

2. Буряк, Н. Ю. Лингвокультурология как наука о взаимодействии языка и культуры / Н. Ю. Буряк // Русский язык и ономастика в поликультурном образовательном пространстве Юга России и Северного Кавказа: проблемы и перспективы : Сборник материалов XI Международной научной конференции, посвященной памяти Заслуженного деятеля науки Адыгеи и Кубани, профессора Розы Юсуфовны Намитоковой, Майкоп, 20–23 декабря 2017 года. – Майкоп: Магарин О.Г., 2017. – С. 120–124.
3. Скнарев, Д. С. Интернет как цифровой инструмент создания и продвижения уникального контента туристического бренда / Д. С. Скнарев, М. М. Русакова // Филологические науки. Научные доклады высшей школы. – 2024. – № S6. – С. 177–186.
4. Hofstede, G.. Culture and Organization: Software of the Mind. Part II. London: McGraw-Hill, 2010. Pp. 53–296.
5. TikTok [Электронный ресурс] // TikTok: сайт. URL: <https://www.tiktok.com/ru-RU/> (дата обращения: 15.04.2025).
6. YouTube Shorts [Электронный ресурс] // YouTube Shorts: сайт. URL: <https://www.youtube.com/@YouTube/shorts> (дата обращения: 15.04.2025).

СОДЕРЖАНИЕ

ПЛЕНАРНОЕ ЗАСЕДАНИЕ

<i>Бредихин С. Н.</i> Конвергенция кодов в виртуальных дискурсивных формах	3
<i>Горожанов А. И., Гусейнова И. А.</i> Создание дидактических материалов для обучения академическому письму на основе корпусных данных	9
<i>Евдокимова В. В., Скрелин П. А., Евграфова К. В.</i> Фонетические характеристики психофизиологического состояния человека	14
<i>Сулейманова О. А.</i> Прагматика и дискурсивная лингвистика: принцип дополнительности?	21
<i>Тивьяева И. В.</i> Возможности больших языковых моделей в развитии исследовательской компетенции студентов-переводчиков	25
<i>Ширяева Т. А.</i> Корпусный анализ ценностных концептов в англоязычном IT-дискурсе	33

Секция 1.

КОРПУСНЫЕ ИССЛЕДОВАНИЯ ЯЗЫКА И РЕЧИ

<i>Алексеева М. Е.</i> Значение прагматического контекста и референциального статуса в оформлении прямого дополнения с квантификатором <i>kol</i> : корпусный анализ данных современного иврита	39
<i>Богданова Н. А.</i> Семантическая просодия немецких глаголов со значением причинности	47
<i>Будникова Е. И.</i> Семасиологические и ономасиологические принципы отбора лексического материала для учебных целей: словарь vs. корпус	57
<i>Ефимова А. Д.</i> Метафоризация и эвфемизация концептуального заимствования «инвестиция» на основе анализа национального корпуса русского языка	63
<i>Ибадова Нигяр Салех кызы</i> Корпусная лингвистика: современный подход к изучению языка	69
<i>Кирина М. А., Лукьянчикова А. С.</i> «Звезды рекомендуют весам пить сливовое вино»: исследование астрологического дискурса на основе распределений частотной лексики и сентимент-анализа	74

<i>Мальмаков В. А.</i> Корпусный анализ политической риторики в речах Дональда Трампа и Джо Байдена	86
<i>Мацегора М. В.</i> Гипербола и парадокс как когнитивные механизмы переосмысления концепта <i>humanity</i> в дискурсе политических и гуманитарных организаций (на материале современного английского языка)	93
<i>Мустафаева Сахила Багхир</i> Дискурс и корпусные исследования в современной лингвистике.....	101
<i>Овсейчик Ю. В.</i> Статистические паттерны сочетаемости связующих элементов сочинительной конструкции в современном французском языке	107
<i>Осьмак Н. А., Коваль И. А.</i> Корпусные технологии в исследовании и преподавании финского невербального поведения	114
<i>Серова А. С.</i> Комбинаторика знака 给 ‘давать’ с иероглифами семантического типа «связи и отношения»: опыт корпусного исследования.....	122
<i>Яковлева В. В.</i> Лингвосемантическая структуризация глаголов <i>schwimmen</i> и <i>плыть</i> в немецком и русском языках	130
<i>Nuriyeva Saadat Tapdıq</i> The Syntax of Double Modality in Standard English: a Corpus-Based Study	138

Секция 2.

ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РЕШЕНИЯ ЛИНГВИСТИЧЕСКИХ ЗАДАЧ

<i>Астапенко Е. В., Митрофанова О. А.</i> Лексико-семантические отношения в корпусах и языковых моделях для белорусского языка.....	147
<i>Безменова Е. С.</i> Искусственный интеллект для интерпретации канонических менасивных высказываний.....	154
<i>Василевская В. М.</i> Оценка качества русскоязычных автоматических рефератов с применением метрик на базе n-грамм (на материале научных текстов из области ИКТ).....	162
<i>Катермина В. В., Гнедаш А. А.</i> Опыт анализа больших лингвистических данных посредством ИИ (на примере исследования российских ВК-сообществ в сфере семьи и родительства)	170
<i>Красовская Ю. Ю.</i> Эффективность синтаксически-обогащенного сентимент-анализа для русскоязычных и англоязычных тревел-блогов	175

<i>Липатова Д. А.</i> Сравнение больших языковых моделей и марковских цепей для генерации анекдотов на русском языке	182
<i>Медведева Н. Г., Яскевич В. В.</i> Использование айтрекинга на основе искусственного интеллекта в исследовании восприятия аккомодационного варьирования английских и русских согласных	191
<i>Михалькова Н. В.</i> Использование методов математической статистики при отборе репрезентативной части массива данных в лингвистических исследованиях.....	200
<i>Мичугина С. В.</i> Возможности ИИ для исследования семантики терминов цвета	208
<i>Николаева М. Н., Овсянникова М. А.</i> Инструменты искусственного интеллекта для корпусного анализа идиом с колоронимом <i>green</i> : сравнительный аспект.....	215
<i>Прокофьева Т. В., Белоглазова Е. В.</i> Анализ тональности руссоцентрического дискурса.....	220
<i>Рябченко Н. А., Малышева О. П.</i> Методологические сдвиги в прикладной лингвистике под влиянием искусственного интеллекта	227
<i>Сяо И.</i> Структурно-семантические модели образования китайских идеограмм, именующих абстрактные понятия: к разработке базы знаний для искусственного интеллекта	234
<i>Тульцева А. Э.</i> Deepseek в немецком медиапространстве: образы, нарративы и вызовы внедрения	242
<i>Царева В. О.</i> О возможностях генерации юмора системами искусственного интеллекта	249
<i>Шеховская Ю. А.</i> Искусственный интеллект и соматические метафоры в политическом дискурсе	256
<i>Kimiaki Tokumaru</i> Restoring scientific memory: regenerative reception of concepts via large language models and neuro-logical integration.....	261

Секция 3. ЦИФРОВАЯ ТРАНСФОРМАЦИЯ ЯЗЫКОВОГО ОБРАЗОВАНИЯ

<i>Аносова Н. Э., Рубцова А. В.</i> Обучение профессионально-ориентированному английскому языку цифровых лингвистов в условиях цифровизации	268
<i>Быканова Н. Н.</i> Обучение умениям рефлексии и компрессии иноязычного текста в письменной речи студентов-программистов с использованием искусственного интеллекта	274

<i>Головач Е. И., Кохан М. Н.</i> Использование искусственного интеллекта в обучении иностранному языку	281
<i>Захарова Е. В., Куварина Е. Н.</i> Интеграция средств дополненной реальности в методику преподавания РКИ: модели и практики.....	287
<i>Зуевская Е. В., Шинкаренко Д. А.</i> Обучение письменной речи на продвинутом этапе с использованием технологий искусственного интеллекта (немецкий язык).....	293
<i>Иваненко Л. А., Шавловская Д. В., Сыч С. А.</i> Использование ИИ в обучении английскому языку в средней школе: анализ диалогового взаимодействия	299
<i>Ковальчук А. М., Кудырко И. В., Трубочкина И. А.</i> Роль электронных образовательных ресурсов в организации воспитательной работы посредством иноязычного образования	307
<i>Костенко С. В.</i> Особенности применения мнемотехник для развития лексических навыков в условиях цифровой среды	316
<i>Кошкарлова Н. Н.</i> Корпусные технологии в преподавании РКИ в странах БРИКС	322
<i>Крылова Е. А., Одинокая М. А.</i> Развитие иноязычных лексических навыков с использованием Spaced Repetition System и цифровых инструментов.....	329
<i>Лебедева Д. В.</i> Аутентичность сгенерированных учебных текстов в обучении иностранному языку	338
<i>Листратова О. К.</i> Прикладное прагматичне забезпечення спеціального призначення як засвідчення оптимізації засвоєння англійської поліграфічної лексики.....	344
<i>Мижевич О. М.</i> Психологические подходы стимулирования учебной мотивации студентов в цифровой образовательной среде	349
<i>Петрова М. Г.</i> Нейрокогнитивные основы адаптивного обучения с ИИ	355
<i>Разорвина А. С., Куварина Е. Н.</i> Искусственный интеллект в образовательных целях: мотивационные, этические и когнитивные аспекты.....	365
<i>Руденко Т. Н.</i> Интеграция цифровых технологий для развития межкультурной компетентности	371
<i>Савчик К. И.</i> Фокус на преимуществах и недостатках цифрового обучения	379

<i>Сироткина Т. А.</i> Компаративный анализ текстов на занятиях по русскому языку как иностранному	384
<i>Тройникова Е. В.</i> Особенности цифровой трансформации межкультурного лингвообразования	389
<i>Шпаковская И. Н.</i> Цифровые технологии как средство оптимизации процесса овладения грамматическим материалом студентами неязыковых вузов	396

Секция 4. МУЛЬТИМОДАЛЬНАЯ КОММУНИКАЦИЯ В ВИРТУАЛЬНОМ ПРОСТРАНСТВЕ

<i>Долженкова О. М.</i> Корреляция мультимодальных элементов и лингвокогнитивных схем и матриц хронотопа в англоязычной новостной экономической статье	402
<i>Евдокимова В. В., Максимова М. Р.</i> Фонетические и мимические характеристики речи при когнитивной нагрузке	409
<i>Зиновьева А. Ю.</i> Эмодзи в анализе тональности (на материале отзывов о медцентрах на русском языке)	423
<i>Керимова Самая Алифхан гызы</i> Интернет-мемы в цифровом пространстве: возможности и перспективы	432
<i>Кустова О. Ю., Шарешик Д. Ю.</i> Лингвокультурная специфика проявления эмоций в поликодовых текстах в контексте цифровой коммуникации	437
<i>Марченко Т. В.</i> Поликодовая дискурсивная репрезентация дифференциальных признаков и атрибутов прецедентных феноменов в медиапространстве	444
<i>Обухова Н. В.</i> Межкультурная медиакоммуникация в цифровую эпоху: сопоставление англоязычных, русскоязычных и узбекоязычных медийных фреймов	449
<i>Силанов Е. А.</i> Медиастилистический анализ цифровых игр как речевого пространства (на примере российских игр «Смута», «Atomic Heart» и «Black Book»)	459
<i>Хрущева О. А.</i> Синкретизм полимодальных текстов в современной интернет-коммуникации	466

Шишкин Б. А. Опыт тематического моделирования и кластеризации
в исследовании виртуального медиадискурса 471

Яскевич Д. М. Особенности сетевого дискурса во франкоязычном
интернет-пространстве 478

Секция 5.
ЦИФРОВЫЕ ТЕХНОЛОГИИ
В ПЕРЕВОДЕ И ОБУЧЕНИИ ПЕРЕВОДУ

Беклемешева Н. Н. Русские деепричастия и способы их передачи
на английский язык: профессиональный перевод vs машинный
(на материале художественной литературы)..... 489

Берендяев М. В., Коканова Е. С. Перспективы дальнейшего развития
классических САТ-инструментов для перевода по мнению
профессионального сообщества 500

Волкова И. Д., Новожилова А. А., Митягина В. А., Никитенко Т. В.
Специфика предредактирования презентационных текстов
для машинного перевода (на материале английского
и немецкого языков)..... 506

Епимахова А. С., Коканова Е. С. Автоматическая оценка качества
машинного перевода в языковой паре с малоресурсным языком..... 517

Мартынова Т. К. Машинный перевод и лексикографическая поддержка:
сравнительный подход в условиях цифровой трансформации языкового
образования 528

Разумова А. В. Возможности использования искусственного интеллекта
при обучении студентов переводу..... 541

Салахиева А. Ю. Использование систем нейронного машинного перевода
для перевода комического (на примере анимационного телесериала
«Inside Job») 548

Чернова И. В., Динаева К. М. Искусственный интеллект
в переводе художественных текстов 562

Чуешкова Э. И. Проблемы эквивалентности медицинских терминов
в нейронном машинном переводе (на примере *Google Translate*
и *Yandex Translate*) 568

Секция 6. МЕЖДИСЦИПЛИНАРНЫЕ ИССЛЕДОВАНИЯ НА ОСНОВЕ АНАЛИЗА ТЕКСТОВЫХ ДАННЫХ

<i>Варановіч В. В., Гецэвіч Ю. С., Супрунчук М. В.</i> Аўтаматызаваная апрацоўка беларускай мовы праз камп'ютарна-лінгвістычны модуль NOOJ	577
<i>Власова Д. В.</i> Психолінгвістические основы изучения англоязычного речевого этикета в современных корпусных исследованиях	586
<i>Евчик Н. С., Меньшенина И. А.</i> Научный дискурс и технологии смысловой интерпретации в условиях цифровой трансформации: от языкового барьера к профессиональному взаимопониманию	592
<i>Зиневич Н. В., Ермилина А. А.</i> Лингвостилистический анализ как инструмент разграничения авторских и роботизированных научных статей	600
<i>Каличева В. М.</i> Картография чувств в киносценарии: построение эмоциональных арок персонажей мелодрам методами машинного обучения.....	610
<i>Кочкурова А. Е.</i> Сентимент-анализ как инструмент исследования музыкального текста (на материале песен группы «Мельница»).....	625
<i>Ли Сюэсинь</i> Воплощение мифообразов в сетературе Китая	639
<i>Полунина Л. Н.</i> Авторство студенческого академического текста: проблема идентификации.....	645
<i>Сафарли Лейла Шамиль гызы</i> Лингвокультурология медиатекста: теоретические основания и аналитические перспективы	653
<i>Свистун Т. И., Русанов А. А.</i> Структурно-языковая специфика статьи-листика	661
<i>Солопова О. А.</i> Макроструктура полиязычного словаря «БРИКС: дискурсивные метафоры»	669
<i>Тарасенко Е. А., Петренко Е. В., Кудрейко Т. В.</i> Проблема авторства и правообладания в ИИ-сгенерированных текстах	677
<i>Холова З. Д.</i> Психолінгвістический анализ лингвокультурных особенностей коротких видеороликов (шортсов) в русскоязычном и испаноязычном цифровом дискурсе	683

Научное издание

ЯЗЫК В ЭПОХУ ЦИФРОВЫХ ТРАНСФОРМАЦИЙ И РАЗВИТИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Сборник научных статей
по итогам II Международной научной конференции
Минск, 23–24 октября 2025 г.

Ответственный за выпуск *А. А. Соловей*

Редакторы: *Ю. И. Варакса, И. В. Нестеренко, Т. М. Нехай*
Компьютерная верстка *Т. С. Соловьевой*

Издатель и полиграфическое исполнение: учреждение образования «Белорусский государственный университет иностранных языков». Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий от 02.06.2014 г. № 1/337. ЛП № 38200000064344 от 17.09.2025 г.
Адрес: ул. Захарова, 21, 220034, г. Минск.