

## ПРОБЛЕМЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ

УДК 004.81

**Крапивин Юрий Борисович**

кандидат технических наук,  
доцент кафедры теоретической  
и прикладной лингвистики  
Белорусский государственный университет  
иностранных языков  
г. Минск, Беларусь

**Yury Krapivin**

PhD in Technical Sciences,  
Associate Professor of the Department  
of Theoretical and Applied Linguistics  
Belarusian State University of Foreign Languages  
Minsk, Belarus  
ybox@list.ru

### ПРОЕКТ ИНСТРУМЕНТАЛЬНО-ПРОГРАММНОГО КОМПЛЕКСА ДЛЯ ПОДГОТОВКИ ДИДАКТИЧЕСКИХ МАТЕРИАЛОВ ПО АНГЛИЙСКОМУ ЯЗЫКУ

В статье сформулированы основные задачи, требующие решения в рамках предложенных и разрабатываемых технологий и инструментально-программного комплекса для подготовки дидактических материалов по английскому языку, закладывающих основу духовно-нравственного и патриотического воспитания студентов и необходимых для организации учебного процесса в учреждении образования «Белорусский государственный университет иностранных языков».

*Ключевые слова: нейронная сеть; большая языковая модель; текстовый корпус; естественный язык; база знаний.*

### THE PROTOTYPE OF AUTOMATED SYSTEM FOR PREPARING DIDACTIC MATERIALS IN ENGLISH

The article presents the main tasks need to be solved within the proposed and developed technology and automated system for preparation of didactic materials to study English, laying the foundation for spiritual, moral and patriotic education of students, and necessary for the organization of the educational process in Belarusian State University of Foreign Languages.

*Key words: neural network; large language model; text corpora; natural language; knowledge base.*

Актуальность применения методов искусственного интеллекта (ИИ) в образовательном процессе очевидна и обусловлена стремительно протекающими этапами цифровизации. Так, Стратегия развития системы образования Республики Беларусь до 2035 года, утвержденная решением коллегии Министерства образования Республики Беларусь от 12.12.2024 № 19.33, предусматривает цифровизацию процессов в системе образования, обозначая цель развития – совершенствование национальной системы образования на основе развивающихся цифровых технологий, подготовку обучающихся к жизни в цифровом обществе, что предполагает и способ достижения цели – внедрение средств автоматизации управления образовательным процессом в учреждениях образования, опираясь на все более активное использование

информационно-коммуникационных, облачных технологий и технологии больших данных. Приоритетными направлениями научной, научно-технической и инновационной деятельности в сфере образования является выполнение экспериментальных проектов, фундаментальных или прикладных научных исследований в сфере образования, обеспечивающих обновление содержания, внедрение инновационных методов и технологий обучения и воспитания, а в сфере молодежной политики – связанных с разработкой и внедрением в практику новых форм и методов работы по патриотическому воспитанию молодежи, привитию уважения к государственности, бережного отношения к культурному наследию и традициям белорусского народа, формированию нравственных качеств и правовой культуры [1; 2; 3].

### **Проект инструментально-программного комплекса**

Самым надежным источником знаний о языке, а также «полигоном», обеспечивающим чистоту лингвистических экспериментов, включая тестирование, оценки естественного языка (ЕЯ) и т.п., является именно сам язык в том виде, в каком он используется в устной и письменной речи, т.е. в виде некоторого определенным образом подобранного конечного множества текстов. Такие тексты принято называть исходным корпусом текстов (ИКТ) заданного ЕЯ, полагая, что они находятся в памяти компьютера. Достоинством обычной формы представления текста является то, что в основе его лежит слово как наиболее естественная единица языка, которая структурируется как в глубину – на морфемы, фонемы, слоги, так и вовне как член словообразовательных, синтаксических и семантических конструкций. Очевидно, что исходный корпус текстов сам по себе не может обеспечить решения всего аспекта задач, связанных с получением количественных оценок языка и испытанием лингвистических гипотез, особенно если речь заходит о более высоких, чем морфологический, уровнях глубины анализа естественного языка. ИКТ является лишь основой создания соответствующих средств в виде совокупности так называемого корпуса виртуальных аннотированных текстов и инструментальных средств доступа, извлечения, анализа и т.п. естественно-языковой информации. Полезность корпуса возрастает, когда он аннотируется, т.е. каждое слово в нем снабжается лексико-грамматическим, синтаксическим или семантическим кодом, в зависимости от уровня обработки текста.

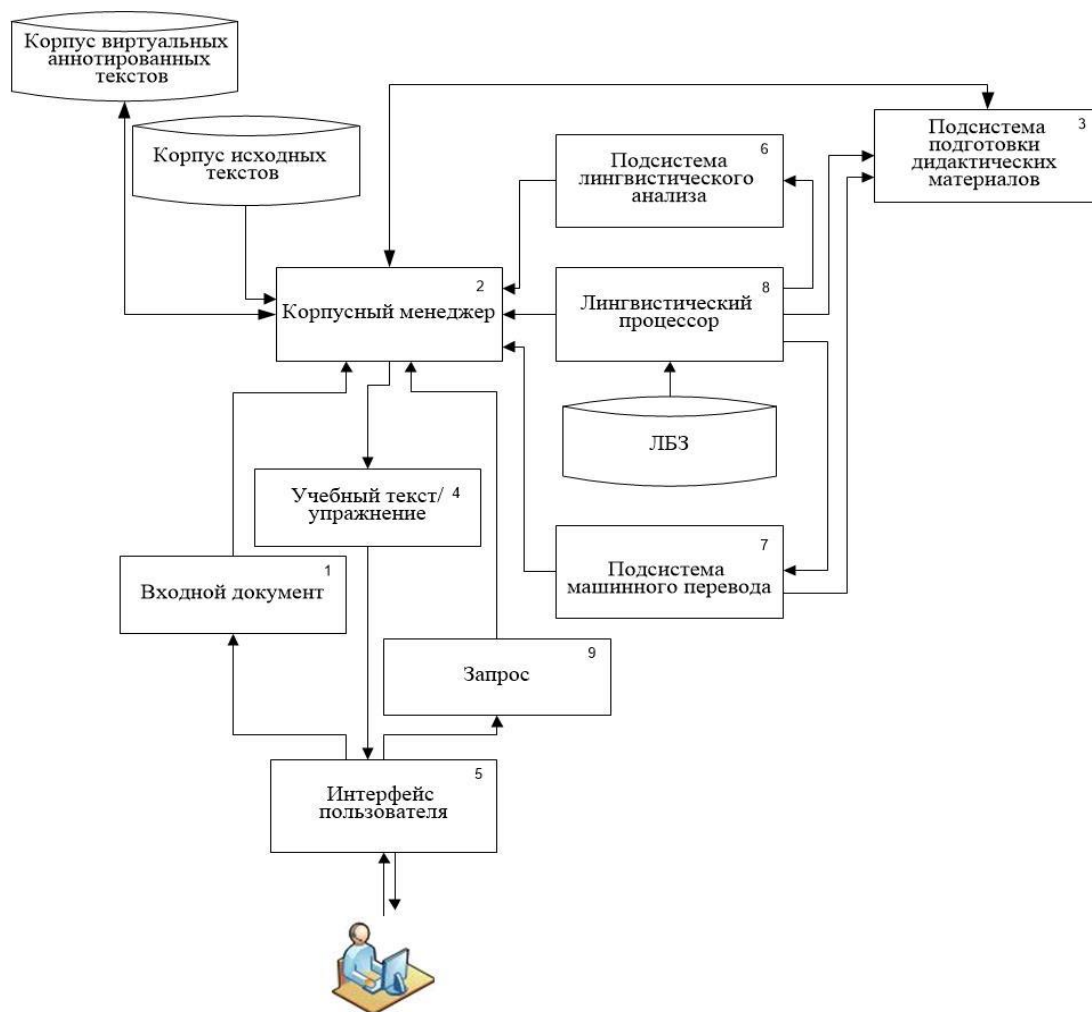
Еще одной важной задачей является подготовка учебных текстов, текстов, которые могут выступать основой для создания разнообразных упражнений и заданий на отработку практических навыков, что достаточно востребовано при реализации программ обучения языковым дисциплинам при овладении иностранными языками и переводу и может основываться на применении таких виртуальных корпусов. Выбор в пользу учебных текстов, содержащих сведения об истории Республики Беларусь, подвиге белорусского народа в период Великой Отечественной войны, трудовых буднях,

спортивных достижениях, современных достижениях в экономике, науке и технике, которые могут быть получены на основании ИКТ, позволяет в дальнейшем рассматривать аннотированный текст не только в качестве хранилища лингвистической информации и изучать, какие лексические единицы наиболее часто используются в определенных типах текстов, какие грамматические конструкции наиболее распространены, но и влиять на восприятие информации, закладывая основу духовно-нравственного и патриотического воспитания учащихся при условии применения таких корпусов текстов в качестве основы для разработки дидактических материалов, в том числе по языковым дисциплинам, преподаваемым в учреждениях общего среднего и высшего образования. Последнее становится особенно актуальным ввиду увеличения масштабов применения цифровых систем и сервисов [4], а также сервисов на базе технологий генеративного искусственного интеллекта, таких как ChatGPT [5], DeepSeek [6], Yandex.GPT [7] и др., применяемых преподавателями учреждений высшего и среднего образования с целью автоматизации и облегчения процессов подготовки дидактических и учебно-программных материалов [8; 9; 10]. Однако подобная практика связана с определенными рисками неявного и непреднамеренного навязывания нежелательной и нехарактерной для молодого человека системы ценностных ориентиров, предубеждений и культурных норм, которые могут транслироваться, используемыми сервисами, через автоматически генерируемые тексты (в широком смысле). Указанный феномен является логическим следствием рисков более общего характера, связанных с применением цифровых инструментов, представленных в работах [11; 12; 13; 14], а также вопросов «качества» используемых текстовых данных, составляющих основу обучения больших языковых моделей [15; 16].

С учетом вышеизложенного разработка технологии и основанного на ней инструментально-программного комплекса для подготовки дидактических материалов по английскому языку, закладывающих основу духовно-нравственного и патриотического воспитания студентов и необходимых для организации учебного процесса на факультете английского языка учреждения образования «Белорусский государственный университет иностранных языков» в соответствии с учебно-программной документацией преподавания языковых дисциплин, видится актуальным и своевременным [3].

Структурно-функциональная схема реализуемого инструментально-программного комплекса представлена на рисунке. Предполагается, что на основании анализа доступных лингвистических ресурсов (полнотекстовых баз данных, тематических корпусов текстов, сетевых электронных ресурсов удаленного доступа и т.д.) выполнен отбор текстов, закладывающих основу духовно-нравственного и патриотического воспитания, и сформирован корпус исходных текстов – *двуязычный русско-английский корпус исходных текстов*. Входной информацией для системы являются Входной документ (блок 1), представленный в одном из наиболее распространенных форматов: TXT, RTF,

DOC, DOCX, PDF, HTML, поступающий на вход Корпусного менеджера (блок 2) для автоматической лингвистической обработки, используя функциональность Лингвистического процессора (блок 8) и Подсистемы лингвистического анализа (блок 6) и, если это необходимо, Подсистемы машинного перевода, с целью ускорения процесса пополнения корпуса исходных текстов и формирования основы построения корпуса виртуальных аннотированных текстов по запросу пользователя (блок 9).



Структурно-функциональная схема  
инструментально-программного комплекса

Корпусный менеджер также использует функциональность подсистемы Лингвистического анализа текстов для выполнения операций навигации, поиска, фильтрации, отбора по заданному условию наборов лингвистических сведений, самих текстов или их фрагментов на заданных естественных языках, по запросам пользователя, а также задействует возможности Подсистемы подготовки дидактических материалов для автоматического построения Учебного текста/упражнения (блок 4) для овладения заданными языками. Последнее реализуется на основании:

◆ результатов лексического, лексико-грамматического и поверхностно-синтаксического анализа, имеющего своим результатом сформированные Лингвистическим процессором кодовые цепочки тегов, отражающие лингвистическую информацию об отдельных элементах предложения, предложении и тексте в целом, являющихся результатом автоматического лингвистического анализа текста, выбранного пользователем из исходного корпуса текстов, заданного в запросе;

◆ содержащихся в лингвистической базе знаний (ЛБЗ) для поддерживаемых ЕЯ множества дидактических правил – типичных шаблонов, определяющих различные типы упражнений для овладения заданными языками, разрабатываемых лингвистами-экспертами в соответствии с программами учебного процесса, а также классификаторов свойств ЕЯ на уровне морфологии и синтаксиса, необходимых для достоверного представления текстов различных предметных областей;

◆ большой языковой модели, построенной на архитектуре Transformer и предварительно обученной или дообученной на наборах текстов из существующего корпуса исходных текстов.

В случае работы системы более чем с двумя языками может потребоваться организация многоязычного исходного корпуса для заданных естественных языков, а функциональность Подсистемы машинного перевода текстов в рамках заданной пары естественных языков автоматизирует процесс получения переводного эквивалента входного текста и гарантирует возможность унифицированной работы с текстами различных естественных языков и предметных областей пользователю ИПК при постобработке переводного эквивалента с целью улучшения качества перевода или устранения возможных неточностей за счет возможного обращения к тематическим и базовым словарям из ЛБЗ.

Таким образом, реализация инструментально-программного комплекса предполагает решение следующих задач [3].

1. На основании анализа учебно-программной документации по языковым дисциплинам, доступных лингвистических ресурсов (полнотекстовых баз данных, тематических корпусов текстов, сетевых электронных ресурсов удаленного доступа и т.д.) построить двуязычный русско-английский корпус исходных текстов, содержащих сведения об истории Республики Беларусь, подвиге белорусского народа в период Великой Отечественной войны, трудовых буднях, спортивных достижениях, современных достижениях в экономике, науке и технике.

2. Используя достижения в сфере информационных технологий и решения на базе больших языковых моделей, разработать лингвистическое обеспечение инструментально-программного комплекса в виде лингвистической базы знаний для русского и английского языков, содержащую лингвистические сведения об указанных языках на уровне морфологии и синтаксиса, а также шаблоны и систему правил, которые в совокупности определяют различные типы учебных текстов и упражнений для овладения английским языком.

3. На основании ЛБЗ, двуязычного русско-английского корпуса исходных текстов разработать алгоритмическое обеспечение инструментально-программного комплекса как набор алгоритмов и решений, позволяющих автоматизировать типовые операции: навигацию, поиск, отбор по заданному условию наборов лингвистических сведений, самих учебных текстов или их фрагментов на заданных естественных языках, формирование двуязычного русско-английского корпуса виртуальных аннотированных текстов по запросам пользователя – преподавателя-автора образовательных ресурсов, учебников, пособий и т.п.

4. Разработать опытный образец инструментально-программного комплекса, провести его эксплуатацию на базе учреждения образования «Белорусский государственный университет иностранных языков» с анализом перспектив дальнейшего развития и практического использования, подготовить соответствующие эксплуатационные и методические материалы.

Предложенное решение, реализуя автоматизацию процессов проведения лингвистического анализа с целью построения двуязычного русско-английского корпуса исходных текстов и формируемого в соответствии с требованиями пользователя двуязычного русско-английского корпуса виртуальных аннотированных текстов, как основы создания учебных текстов и разработки дидактических материалов по языковым дисциплинам, применяя методы на базе искусственных нейронных сетей и лингвистических баз знаний для русского и английского языков, обеспечивает неотделимость обучения от воспитания, способствуя формированию культуры, целостного развития всех сторон личности студента, сохранению и обогащению национальных ценностей, а также использованию информационных технологий в целях воспитания и социализации обучающихся.

## ЛИТЕРАТУРА

1. Образование и молодежная политика : государственная программа. – 2025. – URL: <https://edu.gov.by/gosudarstvennaya-programma-obrazovanie-i-molodezhnaya-politika/> (дата обращения: 30.11.2025).
2. О грантах Президента Республики Беларусь в сферах науки, образования, здравоохранения, культуры, молодежной политики : Указ Президента Республики Беларусь от 13 сентября 2013 г. № 425. – 2025. – URL: [https://etalonline.by/document/?regnum=p31300425&q\\_id=&ysclid=mfgyu3q1na839008118/](https://etalonline.by/document/?regnum=p31300425&q_id=&ysclid=mfgyu3q1na839008118/) (дата обращения : 17.11.2025).
3. Крапивин, Ю. Б. К вопросу разработки инструментально-программного комплекса для подготовки дидактических материалов по английскому языку / Ю. Б. Крапивин // Физико-математическое образование: традиции, инновации, перспективы : материалы III Междунар. науч.-практ. конф., 23–24 окт. 2025 г. / БГПУ. – Минск, 2025. – Ч. 2. – С. 61–65.

4. *Ермашкевич, Н. С.* Развитие профессиональных компетенций в условиях цифровой экономики / Н. С. Ермашкевич, М. Л. Панявина, З. В. Родионова // *Лидерство и менеджмент.* – 2021. – Т. 8. – № 4. – С. 483–502.
5. ChatGPT. – 2025. – URL: <https://chatgpt.org> (date of access : 17.11.2025).
6. DeepSeek. – 2025. – URL: <https://deep-seek.ai> (date of access : 10.11.2025).
7. YandexGPT5. – 2025. – URL: <https://yandex.cloud/ru/services/yandexgpt> (date of access : 10.11.2025).
8. *Troeglazova, A.* Analysis of digital literacy of teachers on the use of distance learning technologies / A. Troeglazova // *Actual Problems of Education.* – ELRA, 2022. – P. 204–207.
9. *Богданова, Н. А.* Использование инструментов генеративного искусственного интеллекта при обучении произношению на продвинутом этапе / Н. А. Богданова // *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. 27–32.
10. *Горожанов, А. И.* Создание лингвистического корпуса на основе инструментов обработки естественного языка: планирование программных решений / А. И. Горожанов // *Филологические науки. Вопросы теории и практики.* – 2023. – Т. 16, № 5. – С. 1616–1620.
11. Intercultural contact in the digital age: A review of emerging research on digitally mediated acculturation / J. Stuart [et al.] // *Advances.in/psychology.* – 2025. – Vol. 2. – P. 299–315.
12. Using Social Virtual Reality in Teaching Intercultural Communication / K. Knutzen [et al.] // *Technology, Knowledge and Learning.* – 2025. – Vol. 30, № 2. – P. 1167–1187.
13. Use of digital technologies in forming business communication skills and cross-cultural communication of students / E. F. Serdyukova [et al.] // *SHS Web of Conferences.* – 2024. – Vol. 195. – P. 1–8.
14. *Chen, H.* Exploring a Mobile Technology-Driven Model for Intercultural Communication Education / H. Chen, K. Mei // *International Journal of Interactive Mobile Technologies (iJIM).* – 2024. – Vol. 18, № 18. – P. 62–75.
15. Источники ChatGPT, Google AI и Perplexity в 2025: откуда нейросети берут информацию – ADPASS. – 2025. – URL: <https://adpass.ru/istochniki-informacii-nejrosetej/> (дата обращения : 27.11.2025).
16. AI Platform Citation Patterns: How ChatGPT, Google AI Overviews, and Perplexity Source Information [Electronic resource]. – 2025. – Mode of access: <https://www.tryprofound.com/blog/ai-platform-citation-patterns> (date of access: 27.11.2025).

*Поступила в редакцию 05.01.2025*