

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.