

УДК 81

Английский язык в цифровых базах данных: обзор влияния, интеграции и перспектив развития

Павлов Игорь Владимирович	pavig2004@mail.ru
Джанаев Карим Сергеевич	karim-2018@yandex.ru
Тихонов Никита Александрович	nikitatihonov432@mail.ru
Полякова Кира Александровна	kir.polyackowa@yandex.ru

КФ МГТУ им. Н.Э. Баумана, Калуга, Россия

Исследовано влияние английского языка в цифровых базах данных на процессы обучения, исследования и лингвистическую практику. Проанализированы преимущества использования языковых корпусов для развития аналитических навыков, структурирования академических публикаций и применения методологии data-driven learning, а также влияние интернета на трансформацию языка. Рассмотрены языковые, методологические, технические и этические вызовы, возникающие при использовании цифровых ресурсов, и предложены пути их решения через стандартизацию данных, внедрение обучающих модулей и адаптацию технологий под современные условия. Подчеркнута значимость дальнейшего развития цифровых инструментов для повышения качества обучения, научных исследований и глобальной коммуникации на английском языке.

Ключевые слова: английский язык, цифровые базы данных, языковые корпуса, обучение иностранным языкам, академические публикации, фразеология, лингвокультурология, DDL

Цифровые базы данных играют ключевую роль как в обучении, так и в научных исследованиях [1]. Еще в конце XX в. Т. Джонс предложил идею обучения иностранному языку на основе баз данных, где обучающийся становится исследователем, а преподаватель — координатором, обеспечивающим необходимый контекст для анализа языковых явлений. Использование языковых корпусов позволяет студентам развивать аналитические навыки, поскольку они учатся распознавать языковые структуры, искать аналогии и обобщать полученные данные. Работа с аутентичными материалами способствует формированию идиоматической речи, близкой к речи носителей, а также помогает пересмотреть традиционные представления о грамматике, так как изучение «живых» речевых примеров делает процесс более естественным и контекстуальным, снижая влияние родного языка. Важно отметить, что современные базы данных часто снабжаются обучающими модулями с тренировочными упражнениями и тестовыми заданиями, что превращает их в полноценные образовательные ресурсы, способствующие интеграции нового научного знания и развитию профессиональной коммуникации [1].

Академические публикации, являясь важным источником данных, обладают сложной структурой, разделенной на абзацы, заголовки, таблицы, рисунки и уравнения, а также содержат встроенные ссылки, что делает их ценным материалом для обработки естественного языка. Например, S2ORC

(Semantic Scholar Open Research Corpus) включает более 81 млн англоязычных работ с сохраненной структурой текста и подробной библиографической информацией, что способствует глубокому анализу дискурса, построению цитатных графов и междисциплинарным исследованиям [2]. Такой интегрированный подход позволяет эффективно решать задачи классификации, синтаксического анализа и выделения сущностей, значительно расширяя возможности информационного поиска.

Важное место в исследовательской практике занимает изучение фразеологических единиц, отражающих национальный культурный код. Фразеология раскрывает скрытую сторону языка, связанную с восприятием, переработкой и категоризацией языковых единиц, а также демонстрирует, как язык и культура взаимосвязаны. Н.Ф. Алефиренко подчеркивает, что в лексическом и фразеологическом фонде языка находят отражение национальный характер, психический склад народа, его история и культура. Исследования, проведенные Гэ Цзиншэнем и И.А. Куприевой, предлагают методологию определения фразеологических единиц на основе таких критериев, как способность составляющих функционировать самостоятельно, наличие двух и более ударных компонентов, структурное сходство, формирование нового значения, сохранение фиксированного композиционного плана, готовность к употреблению, а также коммуникативная направленность и экспрессивность. Такие устойчивые выражения, как *from head to toe*, демонстрируют, что фразеологизмы способны передавать не только прямое, но и коннотативное значение, что открывает перспективы для межъязыкового сопоставительного анализа с использованием мультязычных корпусов [3].

Методология *data-driven learning* (DDL) стала одним из ключевых направлений в современной педагогике. Эффективное изучение языка требует массового контакта с языковыми материалами, однако исследования показывают, что пассивное чтение недостаточно эффективно, например по оценке Zahar et al., час чтения может привести к усвоению всего двух новых слов [4]. DDL дополняет пассивное знакомство с языком формальным обучением, направленным на активное выделение и анализ языковых элементов с помощью компьютерного программного обеспечения. Такой индуктивный подход позволяет учащимся самостоятельно формулировать гипотезы, анализировать частотность, коллокации и распределение языковых единиц, что способствует глубокому усвоению структуры языка. Технологические инструменты, такие как AntConc, делают возможным визуализацию данных и сравнительный анализ, а гибкость метода позволяет интегрировать DDL в CALL-пакеты, адаптируя обучение под индивидуальные потребности [4].

Интернет и современные цифровые технологии оказали существенное влияние на трансформацию английского языка [5]. Стандартный английский ассоциируется с жесткими нормами, элитизмом и строгими правилами, однако большинство пользователей интернета, как носители, так и изучающие язык — активно применяют неформальные, часто деформированные формы, характерные для сетевого общения. Появление аббревиатур, неологизмов, инициализмов и коллоквиальных выражений становится нормой, что приво-

дит к необходимости пересмотра традиционных взглядов на стандартность языка [6]. Различные платформы, адаптированные под компьютеры, планшеты и смартфоны, влияют на стиль письма, а методологии сбора данных, основанные на анкетировании, полевом наблюдении и анализе социальных сетей, позволяют оценить влияние технологий на изменения в употреблении английского языка.

Кроме того, цифровые корпуса, такие как Британский национальный корпус (BNC) в сочетании с инструментами Sketch Engine, предоставляют возможность для детального анализа лексической семантики, что особенно важно для изучающих английский как иностранный язык [7]. Сравнительный анализ синонимов, таких как *learn* и *acquire*, позволяет выявить различия в коллокационных и коллигационных паттернах, где первое слово традиционно ассоциируется с активным освоением навыков, а второе — с пассивным получением знаний. Результаты таких исследований находят применение в учебном процессе, помогая преподавателям объяснять тонкие нюансы употребления и улучшать языковую компетенцию студентов.

Однако использование цифровых баз данных сопряжено с рядом вызовов. Одной из основных проблем является гетерогенность данных, поскольку академические корпуса собираются из различных источников, что приводит к неоднородности текста и метаданных. Современные инструменты, такие как Sketch Engine и AntConc, позволяют нормализовать и унифицировать данные, а постоянное совершенствование алгоритмов разметки помогает улучшать качество извлеченной информации. Методологические сложности возникают из-за необходимости осознания тонкостей статистического анализа, что может стать препятствием для неподготовленных пользователей. Решением этой проблемы является внедрение обучающих модулей и проведение тренингов, направленных на освоение специальных программ и интерпретацию результатов. Технические ограничения, такие как требование платной подписки и разнообразие используемых устройств, также влияют на эффективность использования баз данных, но разработка открытых ресурсов и кросс-платформенных решений постепенно расширяет доступность информации. Этические и правовые вопросы, связанные со сбором и использованием больших объемов данных, решаются посредством стандартов анонимизации и правовых механизмов, что позволяет сохранять баланс между научной пользой и защитой прав пользователей.

В совокупности, несмотря на существующие трудности, цифровые базы данных оказывают глубокое влияние на изучение английского языка и лингвистические исследования, открывая новые возможности для анализа, обучения и глобальной коммуникации. Интеграция методов *data-driven learning*, совершенствование алгоритмов разметки, использование мощных программных средств для анализа языковых корпусов и адаптация к современным цифровым технологиям способствуют успешному преодолению вызовов, связанных с гетерогенностью данных, методологическими сложностями, техническими ограничениями и правовыми вопросами. Перспективы дальнейшего развития кросс-платформенных решений и расширения доступа к от-

крытым источникам обещают улучшение качества обучения, повышение эффективности научных исследований и создание международного сообщества, способного обмениваться знаниями и поддерживать высокие стандарты владения английским языком.

Таким образом, цифровые базы данных становятся не только мощным инструментом в лингвистических исследованиях, но и важным компонентом современной образовательной практики, способствующим трансформации английского языка в условиях стремительного развития технологий и интернета.

Литература

- [1] Иванов И. Базы данных в лингвистике и языковом образовании: современное состояние и возможности их использования при обучении иностранному языку. *Журнал современных исследований*, 2021, № 12 (3), с. 45–60.
- [2] Lo K., Wang L.L., Neumann M., Kinney R., Weld D.S. S2ORC: The Semantic Scholar Open Research Corpus. Arxiv:1911.02782v3, 2019.
- [3] Петрова Е. Лингвистическая основа организации базы данных. *Журнал теоретической и прикладной лингвистики*, 2020, № 4 (2), с. 75–90.
- [4] Boulton A. Data-driven learning and language pedagogy. *Language, Education and Technology*, 2018, pp. 181–192. https://doi.org/10.1007/978-3-319-02237-6_15
- [5] Mares M. The English Language as Internet Lingua Franca. London, University College London, 2017.
- [6] Srinivas Rao P. The Role of English as a global language. Hofuf, King Faisal University, 2019.
- [7] Yang B. A Corpus-based Comparative Study of Learn and Acquire. *English Language Teaching*, 2016, vol. 9, no. 1, pp. 209–220.

English in digital databases: a review of impact, integration and development prospects

Pavlov Igor Vladimirovich

pavig2004@mail.ru

Dzhanaev Karim Sergeevich

karim-2018@yandex.ru

Tikhonov Nikita Alexandrovich

nikitatikhonov432@mail.ru

Polyakova Kira Alexandrovna

kir.polyackowa@yandex.ru

Kaluga Branch of BMSTU, Kaluga, Russia

The article examines the impact of English in digital databases on learning, research and linguistic practice. It analyses the benefits of using language corpora to develop analytical skills, structure academic publications and apply data-driven learning methodology, as well as the impact of the Internet on language transformation. It considers the linguistic, methodological, technical and ethical challenges that arise when using digital resources and suggests ways to solve them through data standardization, implementation of training modules and adaptation of technologies to modern conditions. The conclusion highlights the importance of further development of digital tools to improve the quality of teaching, research and global communication in English.

Keywords: English language, digital databases, language corpora, foreign language teaching, academic publications, phraseology, linguacultural studies, DDL