

УДК 811.111

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ КОРПУСНОЙ ЛИНГВИСТИКИ В ОБУЧЕНИИ АНГЛИЙСКОМУ ЯЗЫКУ ДЛЯ ПРОГРАММИСТОВ

© Семёнов Р.С.

*Поволжский государственный университет телекоммуникаций и информатики,
г. Самара, Российская Федерация*

e-mail: semenovroma206@gmail.com

Научный руководитель: Дукальская И.В., канд. филол. наук, доцент
*Поволжский государственный университет телекоммуникаций и информатики,
г. Самара, Российская Федерация*

Ключевые слова: информационно-коммуникативные технологии, корпусная лингвистика, электронные корпусы, Data-Driven Learning, IT-английский.

Введение

В данной работе исследуется корпусная лингвистика и ее применение в обучении английскому языку программистов.

Как один из международных языков, английский является главным языком для общения программистов по всей планете. Все термины и понятия, которыми оперируют IT-специалисты этой сферы, имеют англоязычное происхождение. Поэтому знание английского языка для них – это незаменимый навык. Корпусная лингвистика, изучающая методы создания текстовых корпусов, в том числе с применением машинных методов, может сильно облегчить процесс изучения структуры языка, значений слов и языковой окраски.

Корпусная лингвистика изучает как создание, так и применение текстовых корпусов. Текстовый корпус обычно определяют как унифицированный, структурированный и размеченный массив языковых (речевых) данных в электронном виде, предназначенный для определенных филологических и, более широко, гуманитарных изысканий [1, с. 89]. Такие массивы данных могут содержать от сотен тысяч до сотен миллиардов словоупотреблений. Такой объем информации предоставит достаточное количество материала для создания профессионально ориентированных курсов английского языка.

Для изучающих иностранный язык существует тип учебного корпуса – электронный корпус текстов, основной целью которого является их анализ на предмет выявления способов и эффективности освоения изучаемого языка. Такие корпусы используются для лингвистического анализа с целью выявления лексических и синтаксических ошибок при освоении иностранного языка, что поможет в составлении учебного плана, коррекции и персонализации в обучении [1, с. 89].

Основные понятия корпусной лингвистики

Корпусная лингвистика – это раздел лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов с применением компьютерных технологий [2, с. 3]. Впервые лингвистические корпусы появились в 60-е годы, а как отдельный раздел науки корпусная лингвистика сформировалась в первой половине 90-х годов. Корпус текстов является объектом корпусной лингвистики, который представляет собой исходный речевой материал для корпусной лингвистики и для других лингвистических дисциплин и сам является результатом деятельности корпусной лингвистики. Корпусная лингвистика разрабатывает теоретические основы и практические механизмы создания и использования представительных массивов языковых данных, предназначенных для лингвистических исследований в интересах широкого круга пользователей [2, с. 3].

Для работы с текстовыми корпусами существует корпусный менеджер, специализированная поисковая система, включающая программные средства для поиска данных в корпусе, получения статистической информации и предоставления результатов пользователю в удобной форме. Это позволяет построить конкорданс, который является списком всех фиксаций искомой языковой единицы в различных контекстах, со ссылками на источник [2, с. 3].

Существуют устные, письменные и смешанные корпуса. Под устным корпусом понимается структурированная совокупность речевых фрагментов, обеспеченная программным средством доступа к ним. Большинство существующих корпусов относятся к письменным или смешанным, однако в смешанных корпусах письменные тексты заметно преобладают над устными. Корпусы также бывают одноязычными, двуязычными и многоязычными. Стоит обратить внимание, что в профорientированном изучении языка стоит использовать одноязычные и двуязычные корпуса [4, с. 82]. Двуязычные корпуса отлично справляются с сопоставлением оригинала текста с его переводом, что крайне полезно в изучении нового языка, а одноязычные корпуса незаменимы при изучении областей языка, относящихся к специальным группам, например корпуса, содержащие тексты, относящиеся к сфере информационных технологий на английском языке, такие как SKELL (Sketch Engine for Language Learning) [6] и один из подкорпусов COCA (Corpus of Contemporary American English) [7].

Лингвистические корпуса используют для обучения языкам по принципу Data Driven Learning (DDL). Общий принцип обучения DDL основан на изучении и анализе данных, полученных из базы данных, из которых обучающийся делает вывод [3, с. 490]. Для изучения иностранных языков используют корпуса, подбираемые под поставленную задачу, в нашем случае из сферы программирования.

Адаптация текста – это упрощение литературно-художественных и других произведений, с целью приспособления информации для малоподготовленных читателей. [5, с. 231]. Адаптация языкового контента на основе реальных языковых данных используется для создания информационной среды для обучающегося, в которую входят упражнения, диалоги и симуляция различных ситуаций. При этом учитывается такой фактор, как частотность слов, сначала идет изучение слов, наиболее часто употребляемых, затем слов, реже употребляемых. Например, в изучении программирования начинают с таких понятий, как «переменная» (variable), «функция» (Function), а в дальнейшем изучаются более глубокие термины: «хэширование» (Hashing), «многопоточность» (Multithreading) и другие. Еще одним фактором являются коллокации – устойчивые сочетания слов, которые часто используются носителями языка. Например, в IT используют не start a script, а run a script.

Анализ применения корпусной лингвистики

Основная идея применения корпусной лингвистики заключается в использовании корпусов для разных задач, требующих большого объема текстовых данных. Проанализировав доступную информацию с сайтов таких платформ, как Skillbox, Skyeng, English4IT, мы выявили несколько способов применения текстовых корпусов для обучения специалистов.

Один из способов – это составление тестовых заданий с помощью адаптирования текстов из корпуса для изучения и закрепления материала. Слова, с которыми ученик испытывал наибольшие трудности, будут попадаться чаще в различных контекстах. Используя корпусный менеджер, ученик может самостоятельно искать значения проблемных слов.

В обучении английскому уже применяются такие корпуса, как British National Corpus, Oxford English Corpus или American National Corpus.

Корпуса, составленные с помощью сайтов GitHub и Stack Overflow, могут помочь начинающим программистам понять значения терминов, благодаря чтению форумов и обсуждений других IT-специалистов. Не исключено, что корпуса могут использоваться для

хранения готового открытого кода. Это означает, что начинающий программист перед тем, как приступить к исполнению задачи, может найти конкорданс требуемой программы и внимательно изучить все варианты для составления собственного кода.

Заключение

В данной статье рассмотрены возможности применения технологий корпусной лингвистики в обучении английскому языку программистов. Как мы выяснили, использование текстовых корпусов позволяет эффективно структурировать учебный процесс, обеспечивая доступ к верным языковым данным, характерным для IT-сферы.

Корпусная лингвистика предоставляет инструменты для анализа частотности лексики, выявления коллокаций и адаптации учебных материалов, что способствует более глубокому усвоению профессиональной терминологии. Инструменты, такие как SKELL и COCA, а также двуязычные корпуса позволяют сопоставлять оригинальные тексты с их переводами и расширять собственные знания.

Метод Data Driven Learning (DDL) – очень эффективный метод обучения, позволяющий студентам самостоятельно анализировать языковые закономерности на основе реальных данных. Кроме того, интеграция корпусов, составленных из материалов GitHub и Stack Overflow, способствует не только изучению языка, но и знакомству с актуальными профессиональными практиками.

Таким образом, применение корпусной лингвистики в обучении английскому языку программистов открывает новые перспективы для создания персонализированных и профессионально ориентированных учебных программ. Дальнейшие исследования могут быть направлены на разработку специализированных IT-корпусов и совершенствование методов их интеграции в образовательный процесс.

Библиографический список

1. Базарова, Б.Б. Корпусная лингвистика и преподавание иностранных языков / Б.Б. Базарова // Вестник Бурятского государственного университета. – 2015. – № 15. – С. 88–92.
2. Захаров, В.П. Корпусная лингвистика: учебно-методическое пособие / В.П. Захаров. – СПб., 2005. – 48 с.
3. Котюрова, И.А. Применение корпусных технологий в обучении иностранным языкам / И.А. Котюрова // Научный журнал РУДН. – 2021. – С. 487–501.
4. Козлова, Н.В. Лингвистические корпуса: определение основных понятий и типология / Н.В. Козлова // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2013. – Т. 11, № 1. – С. 79–89.
5. Мирошникова, Е.А. Адаптация текстового учебного материала при дифференцированном обучении иностранному языку / Е.А. Мирошникова // Вестник Брянского университета. – 2016. – № 3. – С. 229–234.
6. SKELL – Sketch Engine for Language Learning. – URL: <https://www.sketchengine.eu/skell> (дата обращения: 26.05.2025).
7. Corpus of Contemporary American English (COCA). – URL: <https://www.english-corpora.org/coca> (дата обращения: 26.05.2025).