

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

УДК 378.016:811

DOI 10.18101/978-5-9793-1497-6-4-8

КОРПУСНЫЙ ПОДХОД В ОБУЧЕНИИ ИНОСТРАННОМУ ЯЗЫКУ

© **Базарова БаярмаБаировна**

кандидат филологических наук, доцент,

доцент кафедры английского языка и лингводидактики,

Бурятский государственный университет имени Доржи Банзарова

Россия, 670000, г. Улан-Удэ, ул. Смолина, 24а

E-mail: bbbazarova@mail.ru

Информационные технологии, найдя свое неизбежное применение в сфере образования, стали активно использоваться и в гуманитарном образовании. При обучении иностранным языкам применяются методы корпусной лингвистики, расширяющие возможности как преподавателей, в качестве метода и учебного средства, так и студентов, давая им больше самостоятельности в освоении изучаемого языка. В статье рассматривается корпусный подход к изучению английского языка. Хотя эффективность использования корпусов лингвистами и всеми теми, кто преподает и изучает языки, не вызывает сомнений, корпусы все еще не настолько популярны среди преподавателей и, как следствие, среди студентов, чтобы использовать имеющиеся данные.

Ключевые слова: корпусная лингвистика; корпус; обучение иностранному языку; английский язык.

В обучении иностранным языкам открываются новые возможности с развитием компьютерной лингвистики, которая, в свою очередь, дала толчок для появления корпусной лингвистики. Методологический аппарат корпусной лингвистики является перспективным инструментом в теоретическом и практическом обучении иностранному языку. Корпусная лингвистика – раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов (корпусов текстов) с применением компьютерных технологий [Баранов 2001], [Захаров 2005]. Главной целью корпусной лингвистики считают лингвистическое описание языковой системы, когда речевой материал, полученный в ходе изучения коммуникации людей, отражается в корпусе текстов. Под лингвистическим корпусом понимают совокупность текстов, собранных в соответствии с определёнными принципами, размеченных по определённому стандарту и обеспеченных специализированной поисковой системой.

По мнению В. А. Плунгяна, «Смысл корпусно-ориентированной лингвистики в том, что она позволяет изучать действительно существующие в языке, а не мнимые явления» [Плунгян 2008, 17]. За годы становления и развития корпусной лингвистики, когда уже созданы лингвистические корпуса практически всех

развитых языков, лингвисты склоняются к мысли, что исследования, основанные на корпусе, могут помочь в анализе языковых моделей не только в синхронном срезе, но и в **диахронном**, на всех уровнях языка: морфологические, лексические, лексико-грамматические, фонологические, дискурсивные. В качестве одного из преимуществ обращения к корпусу того или иного языка называют тот факт, что в учебниках до сих пор можно встретить искусственно созданные речевые модели в учебных целях, которые с точки зрения носителей языка не всегда аутентичны. В корпусе же отражены реальные, созданные в естественных речевых – письменных и устных – условиях тексты. Очевидно таким образом, что использование подобного ресурса в обучении, равно как и в самостоятельном изучении иностранного языка, имеет свои преимущества.

Sketch Engine - это удобное и эффективное программное обеспечение (см., например, [Базарова 2018]) для управления корпусом и анализа текста на веб-основе, его целью является, прежде всего, предоставление возможности поиска в больших текстовых массивах. Богатые возможности этого ресурса – сотни справочных и специальных корпусов на разных языках – позволяют исследователю подстраивать его под свои собственные задачи. Отличительные его особенности – возможность построить свой корпус, включая разметку и лемматизацию частей речи, структуру слов и различия структур слов на основе сравнительного анализа словосочетаний, автоматический тезаурус, поиск по конкордансу, поиск словосочетаний, списки слов с частотами и обобщенные данные, n-граммы, извлечение терминологии, диахронический анализ и тенденции, а также параллельное (двухязычное) оборудование. Sketch Engine используется крупными британскими и другими издательствами для создания словарей.

Для начала работы в корпусе следует изучить раздел User guide (путеводитель пользователя), где описываются, какие задачи здесь можно решать. Поиск может осуществляться по представленным параметрам. Используя веб-интерфейс, можно искать слова, фразы, леммы и более сложные структуры. Также можно провести анализ частотности употребления; рассмотреть сочетаемость слов. Кроме того, можно проследить жанрово-стилистические особенности употребления единицы языка, например, использование слов и словосочетаний в публицистических текстах или в художественных текстах. Можно создавать и так называемые настраиваемые списки для любой категории, которая представляет интерес, а затем повторно использовать их в последующих запросах (такие как colors + clothes, или слова, относящиеся к beautiful + woman).

Панель управления (dashboard) показывает, какие инструменты доступны для работы с корпусом. Здесь имеются следующие инструменты: Word Sketch, Word Sketch Difference, Thesaurus, Concordance, Wordlist, N-grams, Keywords. Нужная информация генерируется выбранным корпусом автоматически.

Типичные словосочетания представлены в разделе Word Sketch, где необходимо ввести лемму, то есть исходную форму слова. При наборе слова в поиске, Word Sketch подбирает самые распространённые сочетания, сортируя их по грамматическим отношениям. Слова в колонках расположены, начиная от самого распространённого варианта сочетания с ключевым словом. Кроме того, представлены варианты возможных сочетаний как до ключевого слова, так и после него.

Word Sketch один из самых важных разделов в формировании лексической и грамматической валентности. Данный инструмент может стать основой для создания упражнений на формирование навыков лексической сочетаемости.

Раздел Word Sketch Difference может пролить свет на тонкие различия между, на первый взгляд, синонимичными лексемами. Доступны три варианта сравнения: лемм, когда сравнивается использование двух разных лемм через их коллокаты (сочетаемостные особенности); форм слов, когда сравнивается использование двух разных форм слова одной леммы через их коллокаты; подкорпусов, когда сравнивается использование одной и той же леммы в двух разных подкорпусах одного и того же корпуса через их коллокаты. Каждый корпус можно разделить на более мелкие части, подкорпусы. Подкорпус может использоваться для разделения корпуса по типу (художественная литература, газета), форме (устная, письменная) или временной отнесенности (например, по годам) или по любым другим критериям.

Сравнение коллокатов позволяет лучше понять разницу в использовании и значении слов. Word Sketch Difference делает сравнение более эффективным, автоматически генерируя коллокации, важные при сравнении. Функция Sketch Engine назначает цвет каждому поисковому слову и генерирует их графическое описание. Затем он сравнивает грамматические отношения коллокатов отдельно. Цвет указывает на более типичные коллокации, оттенок цвета указывает на силу коллокации.

Для поиска синонимов можно воспользоваться разделом Thesaurus, где также нужно ввести лемму. Тезаурус в Sketch Engine – это автоматически генерируемый список синонимов или слов, принадлежащих к той же категории, он формирует так называемое семантическое поле. Список составляется на основе контекста, в котором слова появляются в выбранном корпусе. В большинстве корпусов поддерживаются только существительные, прилагательные, глаголы и наречия. Качество тезауруса сильно зависит от того, насколько богат выбранный корпус коллокатами. Это означает, что нужен очень большой корпус. Корпус размером около 100 000 слов – это минимум, необходимый для получения внятного результата. Однако для редких слов требуется гораздо больший корпус, чтобы обеспечить данный результат.

Синонимы определяются автоматически в зависимости от контекста, в котором они встречаются. Это положение основывается на дистрибутивной семантике, суть которой в том, что слова, встречающиеся в одном и том же контексте, имеют сходное значение. В Sketch Engine это означает, что слова, имеющие одинаковые коллокации, имеют и одинаковое значение. Чтобы определить синонимы поискового слова, сравниваются коллокаты всех слов с одинаковой частью речи и выделяются те, с которыми поисковое слово имеет наибольшую долю коллокатов.

Использование корпуса, в первую очередь, целесообразно при обучении лексической сочетаемости, кроме того, грамматическую валентность тоже можно вводить и тренировать с помощью корпусного материала. Так, например, функция CQL (Corpus Query Language) в разделе Конкорданс позволяет изучить лексические и грамматические структуры языка в комплексе.

Одна из главных проблем при освоении лексики – разграничение синонимов и их употребление в соответствии с контекстом. Корпус помогает показать

разницу между двумя, на первый взгляд, синонимичными, но различающимися в употреблении лексическими единицами. Раздел Word Sketch Difference дает представление об использовании лексем в различных контекстах. С помощью данного инструмента можно составить упражнения на выявление семантических различий. Список синонимов может содержать слова, которые могут быть никак не связаны с поисковым словом. Это результат автоматической обработки. Sketch Engine не может определить сходство смысла напрямую, он может только сравнивать коллокации. Если два слова имеют одни и те же словосочетания, они будут определены корпусом как синонимы, даже если значения расходятся. Такие случайные неточности не делают инструмент менее полезным. Чтобы избежать этого, необходимо использовать корпус большего размера.

Concordance предоставляет слово в контексте. Это кладезь материала для лексических упражнений. С помощью конкорданса можно составлять лексические УРУ, брать материал для упражнений на подстановку, давать учащимся слова в контексте на предтекстовом этапе чтения или предоставить им возможность самостоятельно отыскать их. Работа с корпусом может выявить синтаксические связи глаголов (управление глаголов), и впоследствии предотвратить интерференцию у учеников. Конкордансу свойственна лингвистическая наглядность, с помощью которой можно выстроить лексико-грамматический профиль. Вертикальное сканирование конкорданса позволяет выявить характерную сочетаемость и наиболее допустимое грамматическое оформление. Предлагаемые параметры помогают рассортировать строки конкорданса с целью усвоения лексико-грамматических особенностей употребления.

Wordlist генерирует списки частотности употребления по части речи, по словоформам, начальным или конечным частям слова, тэгам, леммам и другим характеристикам. Wordlist может быть установлен настройками на минимальную или максимальную частоту.

Раздел N-grams генерирует списки частотности последовательности тех или иных токенов. Пользователь может отрегулировать длину n-граммы до 2-6 единиц. Данный раздел дает знание о частотности использования выражений длиной от 2 до 6 единиц. Элементы внутри n-граммы могут не иметь никакого отношения между ними, за исключением того факта, что они появляются рядом друг с другом.

Раздел Keywords помогает понять тему целевого корпуса (focus corpus). Анализ происходит на сравнении целевого корпуса с референтным корпусом (reference corpus), то есть тем, который обладает репрезентативностью и стандартизованностью. Keywords - слова (tokens), доминирующие в корпусе, определяющие его основное значение и функцию. Terms – ключевые фразы, как правило, состоящие из одного двух слов, основные коллокации специализированного корпуса. Сам раздел подразделяется на ключевые слова в отдельности и словосочетания. Таким образом, можно составить список ключевых слов того или иного корпуса.

Для осуществления поиска по каждому из представленных разделов, представлены две вкладки на выбор: Basic, в случае простого поиска слова в контексте, без привязки к определённым грамматическим характеристикам, и Advanced, где предоставлен более широкий спектр параметров поиска,

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

например, в разделе Word Sketch это выбор части речи или перевод с помощью русскоязычного корпуса.

Неоценимым преимуществом данного ресурса является возможность создания собственного корпуса. Для этого необходимо выбрать тип корпуса (одноязычный или мультязычный), язык, затем загрузить отобранные тексты (свои или взятые в интернете). Компиляция текстов производится программой автоматически. Создаваемый корпус должен строиться согласно учебным задачам, т. е. лексика и грамматика должны соответствовать коммуникативным задачам, запланированным преподавателем. Тексты должны отражать определенную проблематику и мотивировать обучающихся.

Ресурсы и методы корпуса имеют большой потенциал для улучшения педагогической практики и корпуса могут использоваться различными способами, косвенно для дополнения учебных материалов и справочных работ, или непосредственно в качестве инструментов изучения языка и данных для планирования учебной деятельности.

Литература

1. Базарова Б. Б. Sketch Engine в лингвистических исследованиях // Казанская наука. № 5. 2018. С. 28–30.
2. Баранов А. Н. Корпусная лингвистика // Введение в прикладную лингвистику. М., 2001. С. 112–137.
3. Захаров В. П. Корпусная лингвистика: учебно-методическое пособие. СПб., 2005. 48 с.
4. Плуныян В. А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении, 2008, 7–20.
5. Sketch Engine [Электронный ресурс]. URL: <https://the.sketchengine.co.uk>

CORPUS APPROACH IN TEACHING A FOREIGN LANGUAGE

Bayarma B. Bazarova

Ph.D. of Philological Sciences, Associate Professor,
DorzhiiBanzarov Buryat State University
24a Smolina St., Ulan-Ude 670000, Russia
E-mail: bbbazarova@mail.ru

Information technologies are the key issues in education today. In teaching foreign languages methods of corpus linguistics are successfully used extending teaching techniques and giving students more study tools and independence. Corpus-based approach to learning English is touched upon in the article. Though the efficacy of using corpora by linguists and all those who teach and study languages is of no doubt, corpora are still not so popular among teachers and, as a result, among students to use data available.

Keywords: Corpus Linguistics; corpus; teaching a foreign language; English.