

УДК 811.111

ХАРАКТЕРИСТИКИ ПЕДАГОГИЧЕСКОЙ РАБОТЫ С АНГЛИЙСКИМ НАЦИОНАЛЬНЫМ КОРПУСОМ

© **Ирина Васильевна ДЕРЯБИНА**

Тамбовский государственный университет им. Г.Р. Державина, г. Тамбов,
Российская Федерация, аспирант, кафедра лингвистики и лингводидактики,
e-mail: gleamiris@mail.ru

Рассматриваются направления корпусной лингвистики, расшифровывается понятие «корпус текстов». Дается определение четырех основных признаков корпуса, а также описывается применение лингвистического корпуса в учебном процессе. Показано использование методов прикладной лингвистики в преподавании иностранного языка, анализируется потенциал обучения на базе корпусного языкового материала. Представлен анализ таких методов корпусной лингвистики, как автоматизированное извлечение информации (information retrieval (IR)), обучение на основе данных (data-driven learning), текстовые поиски в крупномасштабных корпусах (конкордансы), методы обработки естественного языка (natural language processing (NLP)).

Ключевые слова: корпусная лингвистика; конкорданс; национальный корпус; коммуникативная компетенция.

Процесс информатизации не мог обойти и систему образования. В настоящее время основной задачей в этой области является повышение качества процесса обучения, а также внедрение современных информационных технологий в образование. Одним из таких примеров является Интернет, незаменимый помощник в изучении иностранного языка.

Направления корпусной лингвистики, а также проекты электронных корпусов текстов, активно развиваются и занимают передовые места в методике обучения иностран-

ным языкам, имея значительный прикладной потенциал. В процессе преподавания иностранных языков одна из основных проблем – это нехватка соответствующих педагогических текстовых материалов и актуальных вокабуляров. Нагрузка преподавателей по ежедневной подготовке новых мотивирующих материалов для занятий велика, именно поэтому в данной сфере весомую помощь может оказать применение методов корпусной лингвистики. Они включают в себя: автоматизированную выборку информации; обучение, базирующееся на данных; поиски

по тексту в крупномасштабных корпусах с использованием методов обработки естественного языка.

Само понятие «корпус текстов» все шире входит в научный оборот лингвистов. Именно на его базе развивается корпусная лингвистика. Под корпусом текстов понимается: «унифицированный, структурированный и размеченный массив языковых (речевых) данных в электронном виде, предназначенный для определенных филологических и, более широко, гуманитарных изысканий» [1]. В корпусе данных можно построить конкорданс для любого слова, т. е. список всех употреблений данного слова в контексте со ссылками на его источник.

В своей работе О.В. Нагель описывает обучение на основе данных, или *data-driven learning (DDL)*, которое в последнее десятилетие получила перспективу развития в организации процесса обучения иностранного языка. Основывается оно на капитальном фактическом доказательстве того, что обучающиеся могут гораздо эффективнее осваивают язык, если в процессе образования поощряется использование следующей модели: «наблюдай – предполагай – экспериментировать» (*observe – hypothesize – experiment model*)» [2]. Это означает, что студенты имеют возможность делать свои собственные выводы по поводу значений слов, фраз, грамматических правил, основываясь на аутентичном языковом материале.

Существуют т. н. естественные языковые средства (*Naturally-occurring Language*), которые занимают чрезвычайно важное место в процессе обучения иностранным языкам, т. к. дают студентам возможность изучать только те предложения, с которыми они столкнутся в реальной ситуации общения на иностранном языке в жизни.

Студенты, обучающиеся на основе пуристических учебников с их традиционными описаниями письменного синтаксиса, порой не подготовлены к восприятию и корректному анализу естественной речи, которая изобилует довольно сложными предложениями. Определение корпуса в корпусной лингвистике строится на основе четырех признаков:

- 1) расположение корпуса на машинном носителе (*machine readable form*);
- 2) стандартизованное представление словесного материала на этом машинном но-

сителе, позволяющем применять стандартные программы его обработки, процедуры отбора (*sampling*);

- 3) конечный размер (*design criteria*);

4) репрезентативность (*representativeness* – это название того набора принципов или требований, на основе которых был организован или составлен корпус).

По мнению В.В. Рыкова, первые три признака (*machine readable form, sampling, design criteria*) определяют готовый набор текстов, который располагается на машинном носителе и представляется в удобном для программной обработки виде. Последний же признак (*representativeness*) – это признак репрезентативности. «Корпус должен с максимальной объективностью представить разнообразие изучаемого явления и в то же время дать объективную картину применения этого явления в речевой практике носителя языка» [3, с. 89-96]. Так как именно это свойство лингвистического корпуса внесет незаменимый вклад в образовательный процесс обучения лексики.

Например, для того чтобы узнать в чем различия в употреблении глагола *spend* со следующими предлогами: *on, in* необходимо:

- 1) зайти на интернет-сайт нужного нам лингвистического корпуса;
- 2) ввести в строку «поиск» необходимое слово или словосочетание и нажать **ENTER**;
- 3) на экране появится результат поиска.

Многочисленное использование корпуса большим количеством пользователей обусловило унифицирование и разметки, и программное обеспечение. Лингвистические и экстралингвистические разметки должны основываться на широко распространенных и общепринятых принципах описания текстов и языковых единиц. Параметры разметки, а также их значения должны соответствовать универсальным научным классификациям. Программное обеспечение, в свою очередь, должно заниматься обработкой типовых запросов и решением типовых задач. Огромное значение уделяется унификации форматов, как их наполнения, так и структуры. Обмен корпусными данными происходит на основе единых форматов представления данных, т. е. единого программного обеспечения. Стандартизация корпусов представляющая собой совместимость типов данных

занимает важное место при сравнении разных корпусов.

Работа пользователей с корпусом осуществляется с помощью специальных программных средств – корпусных менеджеров, которые предоставляют разнообразные пути получения из корпуса нужной информации:

- «– поиск конкретных словоформ;
- поиск словоформ по леммам;
- поиск группы словоформ в виде разрывной или неразрывной синтагмы;
- поиск словоформ по набору морфологических признаков;
- отображение информации о происхождении, типе текста и т. п.;
- вывод результатов поиска с указанием контекста заданной длины;
- получение различных лексико-грамматических статистических данных;
- сохранение отобранных строк конкорданса в отдельном файле на компьютере пользователя и др.» [4, с. 12].

Полученные результаты поиска представлены в виде конкорданса, где искомая единица представлена в ее контекстном окружении и в виде статистических данных. Они, фиксируя частотные характеристики отдельных языковых единиц, или грамем, характеризуют совместную встречаемость нескольких лексических единиц.

Национальный корпус применяется не только для научных исследований лексики и грамматики языка, но и для изучения непрерывных процессов языковых изменений, которые происходят в языке в течение определенного периода времени. Еще одна функция корпуса – предоставление различных лексических, грамматических, акцентологических, исторических справок. На данном этапе развития компьютерные технологии намного упрощают и ускоряют лингвистическую обработку немалых массивов текстов. Возможности использования корпуса велики: это и осуществление контроля уровня усвоенного материала, и применение его для составления лексико-грамматических заданий, тестов, создания проектов

Спектр использования корпусов достаточно обширен. К нему относятся и выявление недостатков в уже существующих материалах для преподавания иностранных языков. Исследователи, имея общую методологию, анализируют сначала типовые структу-

ры в учебниках или словарях, а затем их же на базе корпусов. Сравнивая результаты между двумя наборами, большинство таких исследований дали все основания считать, что между данными в учебниках и использованием речевых образцов носителями огромные различия, это и подтверждается корпусами. Учебники могут или не затрагивать наиболее важные аспекты, или выдвигать на первое место обороты, которые редко используются. В результате в таких работах материалы обучения, не принимающие во внимание реальные контексты словоупотребления, вводят в заблуждение и понижают эффективность обучения языка. Таким образом, мы приходим к мнению о том, что использование корпусов должно как можно чаще практиковаться в частности для разработки методических материалов. По мнению Е.П. Сосниной, зарубежные преподаватели уже всюду используют опыт лингвистических корпусов, т. к. «по сути все основные ошибки студентов существуют из-за того, что они не имели наглядного примера использования того или иного слова, выражения, лишь теоретически знали о строении различных по стилю текстов и т. д.» [5, с. 42-45]. Из этого следует, что обучающимся необходима возможность на собственной практике убедиться в использовании того или иного правила, в чем и помогает применение корпусов для самостоятельной работы.

В учебном процессе корпус можно использовать для определения точного значения слов, т. к. существует немало примеров полисемии. В данной работе мы рассмотрим возможность применения корпуса для изучения управления глаголов, поскольку гораздо легче показывать управление на конкретных, «живых» примерах. По мнению П.В. Сыроева, «формирование лексико-грамматических навыков на основе использования лингвистического корпуса возможно исключительно в рамках проблемного подхода». В отличие от широко распространенного репродуктивного подхода, дидактическая сущность которого заключается в пассивном получении обучающимися информации, проблемный подход позволяет активизировать речемыслительную деятельность учащихся и студентов в процессе формирования языковых навыков и развития речевых умений [6; 7].

В результате реализации проблемного подхода обучающиеся на себе испытывают языковые правила и закономерности, превращаясь в первооткрывателей, поэтому знания, полученные на основе такого обучения, лучше всего и надолго усваиваются.

По мнению О.В. Нагеля, эффективность применения лингвистического корпуса в обучении иностранному языку зависит во многом от уровня сформированности иноязычной коммуникативной компетенции. Это означает, что от него будет зависеть эффективность протекания процесса поиска, анализа полученных результатов в системе лингвистических корпусов. Для формирования коммуникативной компетенции О.В. Нагель предлагает использовать лингвистические корпуса таким образом:

«1) для поиска аутентичных примеров использования определенных лексических единиц, устойчивых словосочетаний, грамматических структур и правил, структуры текста;

2) для анализа функции языковых единиц, принимая во внимание как их форму, так и значение;

3) для интенсификации процесса изучения иностранного языка за счет представления информации о наиболее частотных сочетаниях с данной лексической единицей;

4) для расширения словарного запаса студентов в области профессиональной лексики, при использовании специальных корпусов (Special Corpora)» [2].

В первом случае обучающемуся дается возможность самому анализировать полученные данные используя корпус как универсальный источник, а также корректно употребить изучаемые языковые единицы. Кроме того, студент может максимально приблизить свою речь к речи носителей языка. Во втором случае студент рассматривает употребление слова в контексте, что позволяет сузить разрыв между лексическим и грамматическим материалом. В третьем случае лингвистический корпус используется для выявления информации о частотности употреблений необходимых лексических единиц с другими. Это касается и изучения управления глаголов, где лингвистический корпус дает возможность проследить закономерность употребления глагола с тем или иным предлогом.

Основные характеристики метода анализа на базе корпусов, определяющие его надежность и достоверность, – это:

– является эмпирическим и анализирует реальные словоупотребления в естественной языковой среде;

– использует достаточно большую, репрезентативную подборку текстов;

– активно использует компьютеры и специальные программы-конкордансы для анализа в автоматическом и интерактивном режимах работы;

– базируется на методах статистического и качественного анализа текста;

– является целевым, т. е. должен быть ориентирован на реальное приложение и результаты.

Одной из важных особенностей метода анализа на базе корпусов текстов является исследование не только чисто лингвистических явлений (грамматических или лексических функций слов, их связей с другими лексемами), но и таких явлений, как, например, частотности лексем или грамматических конструкций в тех или иных жанрах, диалектах. Итак, мы можем сделать вывод о том, что корпусные методы уже вовсю используются в мировой практике и зарекомендовали себя как высокоэффективные инновационные дополнения к более традиционным образовательным технологиям. Они междисциплинальны, аутентичны и гибки. Также учащиеся имеют возможность поработать самостоятельно. Корпусные методы профильно-ориентированы, что дает возможность адаптировать их к конкретным задачам и целевым группам.

1. *Захаров В.П.* Поисковые системы Интернета как инструмент лингвистических исследований // *Русский язык в интернете.* Казань, 2003. С. 48-59.
2. *Нагель О.В.* Корпусная лингвистика и ее использование в компьютеризированном языковом обучении // *Язык и культура.* 2008. № 4. С. 53-59.
3. *Рыков В.В.* Прагматически ориентированный корпус текстов // *Тверской лингвистический меридиан.* Тверь, 1999. Вып. 3. С. 89-96.
4. *Захаров В.П.* Корпусная лингвистика. СПб., 2005.
5. *Соснина Е.П.* Перспективы корпусного подхода к исследованию грамматических аспектов иностранного языка // *Вопросы теории*

- языка и методики преподавания иностранных языков: сборник трудов международной научной конференции. Таганрог, 2007. С. 42-45.
6. *Сысоев П.В.* Лингвистический корпус в методике обучения иностранным языкам // *Язык и культура*. 2010. № 1 (9). С. 99-111.
7. *Сысоев П.В.* Лингвистический корпус, корпусная лингвистика и методика обучения иностранным языкам // *Иностранные языки в школе*. 2010. № 5. С. 12-21.

Поступила в редакцию 26.06.2012 г.

UDC 811.111

CHARACTERISTICS OF BRITISH NATIONAL CORPUS TEACHING

Irina Vasilyevna DERYABINA, Tambov State University named after G.R. Derzhavin, Tambov, Russian Federation, Post-graduate Student, Linguistics and Linguodidactics Department, e-mail: gleamiris@mail.ru

The issue of the corpus linguistics studies is considered; the concept of "corpus" is deciphered. The author gives a definition of its four characteristics, describes the linguistic corpus application in the studying process. The use of applied linguistics methods in foreign language teaching is given, the teaching potential on the base of corpus language material is analysed. The analysis of such methods as information retrieval (IR), data-driven learning, natural language processing (NLP) and concordances are presented.

Key words: corpora linguistics; concordance; National corpus; communicative competence.