

ИСПОЛЬЗОВАНИЕ ЛИНГВИСТИЧЕСКОГО КОРПУСА В ОБУЧЕНИИ ИНОСТРАННОМУ ЯЗЫКУ

Т.А. Чернякова

Аннотация. Рассматриваются возможности использования лингвистического корпуса в учебном процессе. Представлена характеристика лингвистического корпуса и программы «конкорданс». Проводится анализ критериев отбора текстов в лингвистический корпус, указываются признаки, которыми он обладает. Поясняются условия, которые необходимо учитывать для применения корпуса в учебном процессе.

Ключевые слова: информационно-коммуникационные технологии; лингвистический корпус; конкорданс; обучение лексики.

На протяжении многих столетий причиной войн между Западом и Востоком являлось непонимание культуры, религии, традиций. Современное цивилизованное общество требует от нас достаточного уровня владения механизмами межкультурной коммуникации.

Процесс информационной глобализации все более заметен в мировом сообществе. Уровень информационно-технологического развития влияет на состояние экономики, качество жизни людей, национальную безопасность и роль государства в мировом сообществе. Процесс информатизации также коснулся и системы образования. В настоящее время основная задача в данной области – это улучшение качества учебного процесса и применение современных информационных технологий для повышения уровня образования. Одним из таких примеров является Интернет. Глобальная сеть сегодня занимает около 50% межличностного общения. А ведь именно оно является целью обучения. Общаясь, человек должен уметь понимать представленную информацию, эффективно и понятно строить свою речь. Не менее важно уметь извлекать нужную информацию из прочитанного или услышанного. Каждая речь, письменная или устная, должна находиться в рамках культурных традиций данного общества. Трудно представить, что деловая статья, написанная без учета этих параметров, может быть воспринята серьезно. Следовательно, успех межкультурного общения напрямую зависит от уровня сформированности иноязычной коммуникативной компетенции [1, 2].

Таким образом, проблема обучения эффективной речи остается актуальной в современном образовании. Существует много подходов к решению этой проблемы. Некоторые из них решают ее с учетом произошедших изменений в обществе и науке. По мнению П.В. Сысоева [3], одним из способов является обучение языку, основанное на традиционной риторической тренировке, или обучение на основе образцовых

тренировочных текстов. По сути, человеку, желающему изучить язык, предлагается корпус традиционных текстов, направленных на усвоение и использование их как образцов для дальнейшего общения.

В связи с тем что изобилие текстов представлено в электронной форме и находится в свободном доступе в Интернете, самым большим корпусом можно считать сам Интернет (Web as Corpus), а средствами доступа к этому корпусу являются поисковые машины, например Google. Корпус – это информационно-справочная система, основанная на собрании текстов на некотором языке в электронной форме. Но тексты в Интернете хаотичны, лингвистически интересный запрос часто сложно или невозможно сформулировать с помощью языка запросов поисковой машины, по результатам поиска нельзя оценить представительность выборки (какие тексты не были выложены в сеть или не были найдены данной поисковой системой) и т.д. [4, 5]. По этой причине были созданы национальные лингвистические корпуса. Корпусная лингвистика использует моно- и многоязычные корпуса текстов, которые нередко аннотированы лингвистически значимой информацией, например о частях речи, леммах, морфологических признаках, синтаксической структуре, семантической интерпретации.

Национальный корпус создается лингвистами (специалистами по так называемой корпусной лингвистике, быстро развивающейся современной области языкознания) для научных исследований и обучения языку. Большинство крупных языков мира уже имеет свои национальные корпуса, различающиеся по полноте и уровню научной обработки текстов. Британский национальный корпус (BNC) является общепризнанным образцом. На него ориентированы многие другие современные корпуса. Национальный корпус имеет две важные особенности.

Во-первых, он характеризуется представительностью, или сбалансированным составом текстов. Это означает, что корпус содержит по возможности все типы письменных и устных текстов, представленные в данном языке (художественные разных жанров, публицистические, учебные, научные, деловые, разговорные, диалектные и т.п.), и что все вышеуказанные тексты входят в корпус по возможности пропорционально их доле в языке соответствующего периода. Следует иметь в виду, что хорошая представительность достигается только при значительном объеме корпуса (десятки и сотни миллионов словоупотреблений).

Во-вторых, корпус содержит особую дополнительную информацию о свойствах входящих в него текстов (так называемую разметку, или аннотацию). Разметка – это главная характеристика корпуса. Она отличает корпус от простых коллекций (или «библиотек») текстов, в изобилии представленных в Интернете. Для составителей Национального корпуса такие факторы, как увлекательность или полезность книги, ее

высокие художественные или научные достоинства являются важными, но не первостепенными. Национальный корпус, в отличие от электронной библиотеки, – это не собрание «интересных» или «полезных» текстов для изучения языка. А такими могут оказаться, наряду с классическими произведениями художественной литературы, и роман второстепенного писателя, и запись обычного телефонного разговора, и типовой договор аренды и т.п.

Впервые лингвистический корпус был создан в 1960-е гг. в Брауновском университете (США) на машинном носителе. У. Френсис и Г. Кучера спроектировали его как набор из пятисот двухтысячесловных прозаических печатных текстов американского варианта английского языка. Тексты принадлежали пятнадцати наиболее массовым жанрам англоязычной печатной прозы США. Один из авторов, У. Френсис, употребил слово «корпус» в значении «совокупность текстов, считающаяся представительной для данного языка, диалекта или другого подмножества языка, предназначенная для лингвистического анализа». Появление Брауновского лингвистического корпуса в 1963 г. вызвало всеобщий интерес и оживленные дискуссии. Прежде всего, они коснулись принципов отбора текстов и состава потенциально решаемых посредством данного корпуса задач. У. Френсис и Г. Кучера ставили целью представить корпус текстов, отвечающих ясным и четким критериям отбора:

1. Происхождение и состав текста (автор должен был быть урожденным носителем американского варианта английского языка, диалог должен был занимать менее половины объема текста).

2. Синхронизация (включены были тексты, впервые изданные в 1961 г.).

3. Продуманное соотношение численной представленности различных жанров и отбор отдельных текстов при помощи особой вероятностной процедуры.

4. Доступность для компьютерной обработки (специальные пометы для передачи графических особенностей текста и т.п.) [6, 7].

Так было заложено понятие «лингвистический корпус». В корпусной лингвистике принято определение корпуса, основывающееся на четырех признаках. Этими основными содержательными признаками, которые определяют специфику данного своеобразного рода словесного единства, являются:

- 1) расположение корпуса на машинном носителе (*machine readable form*);

- 2) стандартизованное представление словесного материала на этом машинном носителе, 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).

В настоящий момент в Интернете существует множество лингвистических корпусов, которые классифицируются по жанрам и отраслям и языковой принадлежности. Они могут применяться в учебном процессе, например American National Corpus (ANC), The British National Corpus (BNC), The Reuters Corpus, The Bank of English, English Internet Corpus) [4, 5].

В соответствии с выделенными признаками, которыми обладают лингвистические корпуса, В.В. Рыков [8] делает вывод, что первые три признака (*machine readable form, sampling, design criteria*) определяют некоторый набор текстов, расположенный на машинном носителе и представленный в виде, удобном для программной обработки. Последний признак (*representativeness*), определяющий корпус, есть 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 [8].

Именно это свойство лингвистического корпуса можно и нужно применять в образовательном процессе при обучении лексики на старшей ступени. Применение лингвистического корпуса при обучении техники перевода тоже немаловажно. Итак, **лингвистический корпус** – это система текстов, расположенных на машинном источнике и объединенных по различным параметрам (язык, жанр, стиль, отрасль). Программа, которая помогает анализировать большие массивы текстов и выявлять в них закономерность применения лексических единиц, – «**конкорданс**». Конкорданс осуществляет поиск конкретного словосочетания, слова или грамматического явления в корпусе. После этого открывает в новом окне заданное количество фрагментов текстов, содержащих нужные единицы. На основе полученных сведений можно сделать вывод о функциях, употреблении и применении заданной нами единицы в конкретном языковом пространстве.

Например, нужно узнать, в чем различия в употреблении глагола *spend* с предлогами *on, in*. Для этого 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.

Однако для применения лингвистического корпуса в учебном процессе следует учитывать два важнейших условия:

- ИКТ-компетентность (компетенцию в области использования информационно-компьютерных технологий) [9, 10];
- языковую компетентность (уровень владения иностранным языком).

Национальный корпус предназначен в первую очередь для обеспечения научных исследований лексики и грамматики языка, а также тонких, но непрерывных процессов языковых изменений, происходящих в языке на протяжении сравнительно небольших периодов – от одного до двух столетий. Еще одна функция корпуса – предоставление всевозможных справок, относящихся к указанным областям (лексика, грамматика, акцентология, история языка). Современные компьютерные технологии многократно упрощают и ускоряют процедуры лингвистической обработки больших массивов текстов. Раньше исследователь мог лишь просматривать тексты и вручную выписывать из них нужные примеры; эта предварительная (но абсолютно неизбежная) деятельность была очень трудоемкой и не позволяла обрабатывать большие массивы материала. Теперь ограничений на объем анализируемого материала и скорость поиска информации в нем по существу нет, а это означает, что в распоряжении исследователя оказываются колоссальные массивы текстов самого разного типа.

Теперь подлинно научные описания грамматического строя языков, а также авторитетные академические словари – практически все без исключения – должны составляться на основе корпусов этих языков. Учет корпусных данных оказывается крайне желательным (если не строго обязательным) и при многих других более специальных научных исследованиях.

С этой точки зрения основными потребителями национальных корпусов являются, конечно, исследователи-лингвисты самого разного профиля. Однако круг пользователей корпуса вовсе не ограничивается профессиональными исследователями языка. Надежные статистические данные о языке определенной эпохи или определенного автора могут интересовать литературоведов, историков и представителей многих других областей гуманитарного знания. Важное значение национальные корпуса имеют и для преподавания языка в качестве родного или иностранного; все больше учебников и учебных программ в наше время оказываются ориентированы на корпус. Быстро и эффективно проверить с помощью корпуса особенности употребления незнакомого слова или грамматической формы у авторитетных авторов сможет и иностранец, и школьник, и учитель, и журналист, и редактор, и писатель. Таким образом, национальный корпус обращен ко всем, кто в силу профессии, по необходимости или из простой любознательности ищет ответы на вопросы об устройстве и функционировании языка, т.е. фактически к большинству образованных носителей этого языка и ко всем, изучающим его в качестве иностранного.

Особое внимание все же стоит уделить применению лингвистического корпуса непосредственно в учебном процессе. Корпус можно использовать для определения точного значения слов, так как существует немало примеров полисемии. Возможно применение корпуса для изуче-

ния управления глаголов, так как гораздо легче показывать управление на конкретных, «живых» примерах.

Использовать корпус также возможно для осуществления контроля уровня усвоенного материала. Здесь можно применить корпус для составления лексико-грамматических заданий, тестов, создания проектов.

Таким образом, применение лингвистических корпусов при обучении иностранному языку дает возможность оценить употребление различных словообразований и словосочетаний непосредственно носителями языка, т.е. изучать «живой язык». Более того, лингвистический корпус дает возможность отслеживать изменения и преобразования в языке, так как он постоянно обновляется; позволяет обучающимся самим делать выводы. Для преподавателей лингвистический корпус – это новые возможности в методике обучения иностранным языкам.

Литература

1. **Баранов А.Н.** Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема // Русистика сегодня. 1998. № 1–2. С. 179–191.
2. **Баранов А.Н.** Проблема репрезентативности корпуса данных (на примере политической метафоры) // Труды Международного семинара «Диалог – 2001» по компьютерной лингвистике и ее приложениям. Аксаково, 2001.
3. **Сысоев П.В.** Интегративное обучение грамматике: исследование на материале английского языка // Иностранные языки в школе. 2003. № 6. С. 25–31.
4. **Сысоев П.В.** Лингвистический корпус в методике обучения иностранным языкам // Язык и культура. 2010. № 1. С. 99–111.
5. **Сысоев П.В.** Лингвистический корпус, корпусная лингвистика и методика обучения иностранным языкам // Иностранные языки в школе. 2010. № 5. С. 12–21.
6. **Рыков В.В.** Корпус текстов как отражение состояния русского языка // Труды Международного конгресса «Русский язык: исторические судьбы и современность». М. : МГУ, 2001.
7. **Рыков В.В.** Корпусная лингвистика (научно-аналитический обзор) // Социальные и гуманитарные науки: Зарубежная литература. М. : ИНИОН, 1996. № 4. С. 43–51.
8. **Рыков В.В.** Прагматически ориентированный корпус текстов // Диалог-99. М. : Таруса, 1999.
9. **Евстигнеев М.Н.** Структура ИКТ-компетентности учителя иностранного языка // Язык и культура. 2011. № 1 (13). С. 119–125.
10. **Сысоев П.В., Евстигнеев М.Н.** Компетенция учителя иностранного языка в области использования информационно-коммуникационных технологий: определение понятий и компонентный состав // Иностранные языки в школе. 2011. № 6. С. 16–20.

THE USE OF LINGUISTIC CORPUS IN TEACHING A FOREIGN LANGUAGE Chernyakova T.A.

Summary. This article addresses the issue of the use of linguistic corpus in language teaching. It discusses the characteristics of linguistic corpus and concordance, criteria for text selection, characteristic features of the linguistic corpus. In the paper the author explains conditions for the use of linguistic corpus in language teaching.

Key words: information and communication technologies; linguistic corpus; the web concordance; vocabulary.