

УДК 81'272:[372.881.1+81'374'25]

Метлицкая Наталья Анатольевна
старший преподаватель
Белорусский государственный
университет
г. Минск, Беларусь

Natalia Metlitskaya
Senior teacher
Belarusian State University
Minsk, Belarus
magnatusik@gmail.com

ПРИМЕНЕНИЕ КОРПУСНЫХ ТЕХНОЛОГИЙ ДЛЯ РЕШЕНИЯ ЛИНГВИСТИЧЕСКИХ ЗАДАЧ

В статье рассматриваются возможности применения корпусных технологий на современном этапе. Благодаря своим свойствам (репрезентативность, электронный формат, аннотирование, дву-/многоязычие, параллельность) компьютерный корпус текстов является действенным электронным инструментом для решения широкого круга прикладных задач, связанных с языком. В частности, в данной статье представлены примеры использования корпусных технологий в таких сферах, как лексикография, обучение иностранным языкам и переводческая отрасль.

Ключевые слова: электронный корпус текстов; информационные технологии; конкорданс; лексикография; дидактические свойства; методические функции.

APPLICATION OF CORPUS TECHNOLOGIES IN SOLVING LINGUISTIC TASKS

The article focuses on the usage of corpus technologies at the present time. Due to its properties, a text corpus presents an effective tool in solving various practical linguistic tasks. In particular, it is widely used nowadays in such spheres as lexicography, the process of teaching/learning foreign languages and translating. The article provides examples of such a usage.

Key words: electronic text corpus; information technologies; concordance; lexicography; didactic properties; methodological functions.

В последнее время наблюдается возрастающий интерес исследователей к изучению теоретических и практических аспектов использования корпусных технологий для решения широкого круга различных лингвистических задач. Такой интерес не случаен: в исследовании языка во всем многообразии его функций и вариантов исходным моментом являются языковые данные – реальные факты использования языковых единиц в речевой коммуникации. Использование электронных корпусов текстов при решении лингвистической задачи позволяет в наиболее полном объеме получить необходимые данные, поскольку языковые данные разного типа представлены в корпусе в своей естественной контекстной форме. «В этом смысле лингвистический корпус представляет неограниченные возможности, т. к. в процессе обработки текста происходит максимальная опора на контекст» [1, с. 56]. Причем в рамках корпусных исследований выделяют несколько уровней понимания контекста [1, с. 57]: 1) минимальный контекст, в котором реализуются лексические и морфолого-синтаксические явления; 2) текстовый контекст, включающий в себя фрагменты текста вплоть до текста целиком; 3) контекст, предполагающий учет текстов определенного типа (заданного функционального стиля, отобранной коллекции текстов и т. д.).

Компьютерный корпус некоторого языка представляет собой совокупность языковых данных в электронном виде, которая содержит образцы речи, созданные говорящими в определенных ситуациях общения. Поскольку входящие в корпус примеры речепотребления являются уже «состоявшимися» высказываниями, а не искусственно смоделированными предложениями,

анализ этих данных позволяет расширить наше представление о динамике функционирования и взаимодействия языковых средств в составе дискурса. Разработка более совершенных программ компьютерного обеспечения, со своей стороны, открывает перед исследователем возможность автоматического вывода информации как о системных свойствах языка, так и о его дискурсивных особенностях, связанных со спецификой контекстов употребления в функционально-дифференцированной речи. Вышеизложенным можно объяснить разнообразие задач, для решения которых успешно применяются корпусные технологии на современном этапе. Рассмотрим подробнее некоторые из них.

Например, корпусные исследования являются незаменимыми там, где во главу угла ставятся закономерности речи, обусловленные количественными показателями, такими как частотность или встречаемость. Потому одна из главных областей применения рассматриваемых технологий – лексикография и лексикология. Данные, полученные из корпусов текстов, применяются для составления словарей, определения значений многозначных слов, выявления ассоциативных связей слов в тексте, выделения терминов и терминологических словосочетаний и т.п. Уже в конце 60-х гг. XX столетия наряду с возникновением и развитием корпусной лингвистики во многих странах были инициированы корпусные проекты различных масштабов на разных языках и для разнообразных целей. Так, в США и Европе крупнейшие лексикографические дома обратились к созданию корпусов литературного английского языка (Brown Corpus, British National Corpus, Collins Wordbanks Online English Corpus, Cambridge English Corpus, The Longman Learners' Corpus, The Macmillan World English Corpus, CORPORA, Oxford English Corpus и т. п.) для дальнейшего составления и издания словарей [2, с. 17]. В настоящее время ведущие издательства мира продолжают активно использовать электронные корпуса для разработки лексикографических продуктов. Например, авторитетное издательство Oxford University Press базируется на British National Corpus (более 1,5 млрд слов), лексикографические произведения серии Collins основаны на языковом корпусе The Bank of English (более 1 млрд слов), словари серии Лонгман – на Longman Mini Concordances [3, с. 9]. На основе лингвистического корпуса разработан также известный словарь английского языка «Oxford Dictionary of English, 3rd edition», который был издан в 2010 г. [4, с. 262]. В ходе исследований на основе корпуса Oxford English Corpus были выявлены тысячи новых значений слов, отраженных в данном словаре. В свою очередь, частотный словарь английского языка «Word Frequencies in Written and Spoken English» создан на основе корпуса BNC (The British National Corpus) [4, с. 262].

Для лексикологии проведение исследований при поддержке электронных корпусов стало общепринятым стандартом. «На этом фоне перспективными оказываются лексикографические исследования, опирающиеся на данные корпусного анализа и позволяющие обнаружить влияние последнего на структуру словарных статей и выбор иллюстративных примеров. Методы

корпусной лингвистики применяются для анализа парафраз значения словарных единиц с целью оптимизации дефиниций значения в электронных словарях» [5, с. 1].

Таким образом, благодаря корпусным технологиям лексикографы сегодня имеют возможность в короткие сроки проводить лингвистические исследования на основе богатого эмпирического материала и разрабатывать методики для компьютерного анализа текстов [4, с. 262]. Согласно В. П. Захарову, «все современные лингвистические исследования и работы по составлению словарей и грамматик так или иначе ориентированы на использование представительных корпусов текстов» [6, с. 10].

Еще одной сферой, где в последние годы широко используются достижения корпусной лингвистики, является преподавание иностранных языков. Очевидно, что современные информационные технологии активно применяются сегодня в процессе обучения иностранным языкам, расширяя рамки образовательного процесса. Наряду с традиционными источниками (учебники, словари, художественная литература) все более важную роль в процессе обучения играют электронные источники языковых знаний (Интернет, компьютерные программы, онлайн-ресурсы). Такие информационные средства повышают эффективность обучения, помогают развивать исследовательские и творческие навыки обучающегося в рамках личностно-ориентированного подхода, тем самым способствуя достижению главной цели обучения иностранному языку – формированию коммуникативной компетенции. Роль корпусных технологий в данном процессе заметно возрастает. По мнению профессора В. А. Плунгяна, «теперь для овладения языком человеку нужны не две, а три вещи: словарь, грамматика и корпус текстов данного языка. Потому что и словарь, и грамматика, в общем-то, бесполезны вне этого живого пространства, где язык, собственно, и функционирует» [7].

За последние несколько лет в методической литературе появился целый цикл исследований, посвященный обучению посредством лингвистических корпусов лексической и грамматической сторонам иноязычной речи. Во многих работах рассмотрена проблема выявления и изучения коллокаций на основе корпусов текстов. Иные видят особую актуальность использования корпусных технологий при обучении продуктивным видам речевой деятельности, а именно говорению и письму. Как видно, корпус текстов используется в методике преподавания иностранных языков для решения самых разных задач.

В целом анализ работ, посвященных данной проблеме, позволяет выявить два основных направления применения корпусных технологий в процессе обучения иностранному языку. Первое связано с разработкой учебных пособий и обучающих программ на основе электронных корпусов текстов. А. С. Герд и В. П. Захаров отмечают, что «корпуса могут служить источником и инструментом многоаспектных лексикографических работ, использоваться для уточнения существующих грамматик и составления новых» [8, с. 122]. Среди известных учебных пособий на основе корпусных технологий можно выделить грамматическое пособие издательства «Лонгмэн»

«Longman Student Grammar of Spoken and Written English», изданное в 2002 году. Данное учебное пособие составлено на основе корпуса LSWE (Longman Spoken and Written English Corpus). Авторам удалось проанализировать более 40 млн слов из устного и письменного корпуса английского языка; в тексте приводится более 3000 примеров, иллюстрирующих то или иное грамматическое явление [4, с. 262]. На основе корпуса CIC (North America, Cambridge International Corpus), который содержит более 1 млрд слов, было разработано пособие по грамматике устного и письменного английского языка «Cambridge Grammar of English». В свою очередь, диалоги и учебный материал в серии популярных учебников «Touchstone» также основываются на обширном исследовании указанного корпуса, включающего большую базу данных повседневных диалогов. В результате учебник помогает учащимся общаться эффективно даже на начальном уровне [2, с. 19].

Второе направление предполагает прямое использование корпусных технологий в методике обучения иностранным языкам. Одним из первых использование корпусных данных в преподавании иностранного языка предложил Т. Джонс в 1991 году. Впервые лингвистический корпус был представлен как материал обучения иностранному языку, а в качестве инструмента обучения выступала программная оболочка корпуса. Такой подход получил название Data Driven Learning (DDL) – обучение с помощью базы данных [4, с. 263]. Т. Джонс первым стал применять корпусы на уроках грамматики и лексики на занятиях по английскому языку со студентами международного факультета в Бирмингемском университете. Ученый предложил студентам выступить в роли исследователей, которые с помощью компьютерных технологий и корпусов должны были проводить собственные мини-исследования языка. В дальнейшем использование корпусно-ориентированного подхода к преподаванию иностранных языков получило широкое распространение. Это обосновано несколькими причинами.

Во-первых, в рамках данного подхода роли преподавателя и студента кардинально меняются, что полностью соответствует современной образовательной парадигме, где во главе угла находятся идеи компетентностного и личностно-ориентированного подхода в обучении. Основная задача преподавателя состоит в том, чтобы составлять задания на основе корпусов и мотивировать обучающегося проводить собственное исследование, направлять его в поиске собственных методов выполнения заданий. Такие задания могут быть ориентированы на определения значения слов или изучение близких по значению слов, на выявление случаев полисемии, определение значения терминов и т.д. Задания могут включать непосредственное использование конкорданса, поисковых возможностей корпуса и использование результатов корпусной статистики. В свою очередь, основная задача студента – «научиться учиться», быть субъектом своей учебной деятельности. Во время выполнения заданий с помощью корпуса обучающийся проводит свое исследование языка, переходя от общего к частному: выявляет закономерности, приходит к пониманию формы и значения, наблюдая определенное языковое явление на примерах, предложенных в корпусе. Например, благодаря построению конкордансов на основе корпуса обучающийся получает большое количество

примеров употребления слова или грамматического явления в естественном языковом окружении. Грамматические структуры представлены в корпусе в лексических контекстах, а лексические единицы – в грамматических структурах [4, с. 264]. Таким образом, для обучаемого сокращается разрыв между лексикой и грамматикой.

Еще одной причиной, почему электронный корпус получил большую популярность в образовательном процессе, является быстрота и надежность в поиске необходимой информации. Для поиска и классификации информации в электронном корпусе используется программа конкорданс. Согласно Дж. Синклеру, конкорданс – это собрание случаев использования словоформы, каждая из которых употребляется в разном контексте [9, р. 32]. Именно наличие конкорданса определяет суть корпуса как исследовательского аппарата и делает его невероятно ценным ресурсом в обучении иноязычной лексике [10, с. 73]. Благодаря конкордансу, электронные лингвистические корпуса могут использоваться для выявления информации и статистических данных о языковых и речевых единицах. Например, наблюдая контекстное окружение какого-либо слова, можно установить определенные семантические признаки, по которым в дальнейшем можно будет распознать значение данного слова.

Исследователь В. В. Ключихин отмечает, что электронный лингвистический корпус обладает рядом отличительных дидактических свойств и методических функций, которые выделяют его среди остальных информационно-коммуникационных технологий обучения иностранному языку [10, с. 74]. Согласно определению П. В. Сысоева, дидактическими свойствами лингвистического корпуса следует называть «основные характеристики и признаки, отличающие одни информационные технологии от других. Под методическими функциями понимаются внешние проявления данных технологий» [11, с. 122].

Анализ исследований, посвященных данной проблеме, позволяет выявить набор основных дидактических свойств электронного корпуса. Такими являются:

1. Многоязычие. Лингвистический корпус может включать в себя оригинальный текст с его переводом на другой язык или на несколько языков.

2. Возможность поиска конкретной информации. Современный электронный корпус обладает возможностью поиска конкретной информации (запрашиваемое слово в коллокации с выбранной частью речи, поиск случаев употребления только начальной формы глагола и т. п.).

3. Контекстность результатов поиска. Корпус предоставляет результаты поиска слов в контексте их использования.

4. Разнообразие функциональных типов текстов. Корпус содержит тексты различных типов, стилей и жанров.

5. Сортировка результатов поиска. Лингвистический корпус дает возможность систематизации результатов поиска по определенным критериям (частотность употребления слова на определенном промежутке времени, выбор функционального типа текста и т. п.).

6. Актуальность текстовых данных. В современных электронных корпусах происходят постоянные обновления баз данных, что позволяет следить за тенденциями развития иностранного языка.

7. Общедоступность в сети Интернет. Большинство корпусов доступно для любого интернет-пользователя.

Очевидно, что электронный лингвистический корпус обладает большим набором дидактических свойств, благодаря которым он качественно выделяется на фоне других информационно-коммуникационных технологий.

К методическим функциям электронного лингвистического корпуса относятся следующие [10, с.75]:

- на основе корпусов параллельных текстов возможно формирование лексических навыков посредством перевода;
- благодаря контекстности результатов при помощи лингвистического корпуса возможно более точно дать определение слову, а также выделить наиболее часто встречаемые коллокации данного слова;
- в лингвистических корпусах представлены современные лексико-грамматические структуры, которые позволяют отслеживать и определять значения неологизмов;
- лингвистический корпус имеет функцию поиска только конкретно нужного слова или коллокации, что позволяет интенсифицировать процесс обучения;
- с помощью лингвистического корпуса можно найти значение слова в контексте его использования в определенной сфере жизни общества (медицина, бизнес, авиаперевозки и др.);
- лингвистические корпуса располагают выбором различного уровня языковой сложности, что дает вариативность при выборе лексической единицы для употребления в речевой деятельности.

Таким образом, в процессе обучения иностранному языку лингвистический корпус может использоваться для формирования лексико-грамматических навыков, а также социолингвистической и социокультурной компетенций. Корпус помогает трансформировать огромное количество иностранных текстов в компонент содержания обучения, на базе которого возможно создание учебных заданий, направленных на развитие различных языковых и речевых навыков. Все это делает электронный корпус эффективным вспомогательным ресурсом в решении методических задач.

Говоря о сферах применения корпусных технологий на современном этапе, нельзя не отметить переводческую сферу. Такие свойства корпусов текстов, как репрезентативность, пополняемость, электронный формат, дву-/многоязычие, аннотирование (лингвистическая разметка) делают актуальным их использование в деятельности переводчика. Как отмечает Д. Ю. Груздев, «электронные корпуса дают возможность реализовать творческий подход в поиске переводческих решений и являются надежным источником информации» [12, л. 147]. По мнению данного исследователя электронный корпус текстов «представляет собой ресурс, который содержит

эталонны для оценки качества перевода в тех или иных “стандартных” условиях. Нам нужен именно “живой” язык в том виде, в каком он предстает в реальной речевой деятельности, а не объяснение явлений в нем» [12, л. 16]. В данном случае автор акцентирует внимание на том, что в корпусе язык представлен в своем естественном виде, иллюстрируя реальные ситуации своего функционирования. Данный факт подчеркивает и исследователь Т. В. Бусел: «Несомненным достоинством компьютерного корпуса текстов является то, что он дает представление о типовой лексической и синтаксической сочетаемости встречающихся в нем слов, об их окружении и особенностях функционирования в речи, сообщает наиболее полную информацию о нормах словоупотребления в современном языке» [13, с. 409].

Анализ работ, посвященных исследованию потенциальных возможностей использования электронных корпусов в процессе профессионального перевода, позволяет выявить ряд случаев, когда электронный корпус может оказаться полезным для письменного переводчика: при определении левого и/или правого окружения (лексико-грамматической сочетаемости) слова; при выборе из нескольких вариантов лексического эквивалента исходного слова, предлагаемых в разных словарях или встретившихся в Интернете; при проверке правильности решения, выбранного переводчиком; для поиска дополнительной энциклопедической информации по теме; для поиска терминологических дублетов, антонимов, номенклатурных наименований и дефиниций терминов; для расшифровки встретившихся в оригинале сокращений.

По мнению Н. В. Владимова, корпуса являются более удобным и достоверным средством по сравнению со словарями по ряду причин [14]:

1) корпуса текстов – это не застывшие структуры, как традиционные словари, а постоянно пополняющиеся базы данных. Благодаря этому переводчик получает возможность быть в курсе самых последних тенденций развития языка на основе анализа употребления того или иного слова в корпусе;

2) корпус является более репрезентативным источником словоупотребления, чем словарь в силу того, что он намного больше по объему, а информация о слове, которую можно получить из корпуса, более объективна и точна. Работать с корпусом гораздо удобнее, чем со словарем.

В свою очередь, исследователь Д. Ю. Груздев уточняет, что корпус текстов применяется в качестве вспомогательного средства переводчика и не может полностью заменить словари. Согласно ученому, «корпус текстов представляет собой инструментарий, который существенно облегчает работу переводчика на том этапе процесса перевода, когда он выполняет перебор переводческих вариантов, работая по методу проб и ошибок. Суть этого подхода «заключается в последовательном приближении к оптимальному варианту путем перебора нескольких возможных вариантов перевода и отклонения тех из них, которые не соответствуют определенным функциональным критериям» [12, л. 52]. В процедуре перебора вариантов выделяется два основных этапа: (1) генерирование вариантов и (2) их диагностика, тести-

рование. Сначала переводчик выдвигает как можно больше предположений относительно возможности перевода проблемного участка текста. Главную роль на этом этапе играет компетентность переводчика и умение пользоваться словарями, а также другими вспомогательными материалами. После этого проводится тестирование вариантов перевода. Перед переводчиком стоит задача проанализировать все варианты и выбрать только тот, который обеспечивает наиболее адекватный перевод. «Электронный корпус текстов в силу своих свойств, идеально подходит для второго этапа указанного выше метода и не может заменить словари. Предполагается, что переводчик использует корпус в качестве последней инстанции для проверки своего решения» [12, л. 53].

Как видно из вышеизложенного, корпуса текстов являются электронными инструментами, которые дают возможность профессиональным переводчикам повысить скорость, эффективность и качество выполняемых переводов и потому играют все более важную роль в переводческой деятельности.

Таким образом, корпусные технологии представляют собой пример современных информационных технологий, которые в силу своих свойств и возможностей находят широкое применение при решении различных прикладных задач языкового характера.

ЛИТЕРАТУРА

1. Павельева Т. Ю. Изучение коллокаций на основе лингвистических корпусов текстов // Вестник Тамбовского ун-та. Сер. Гуманитарные науки, 2016. Т. 21. Вып. 3-4 (155-156). С. 56–61.

2. Маник С. А. Современная корпусная лексикография: особенности и перспективы развития // Вестник Ивановского гос. ун-та. Сер. Гуманитарные науки, 2016. Вып. 3 (9). С. 17–20.

3. Работалова Е. В. Корпусные технологии в лексикографии (на примере английской терминологии нефтепереработки) // Вестник Московского гос. обл. ун-та. Сер. Лингвистика, 2020. № 5. С. 8–18.

4. Чернявская О. Г. Дидактический потенциал корпусных технологий в преподавании иностранных языков // Ученые записки Крымского федерального ун-та им. В. И. Вернадского. Филологические науки, 2016 г. Т. 2 (68). № 2. Ч. 2. С. 261–266.

5. Колпакова Г. В. Корпусная лингвистика и лексикография // Электронный научно-образовательный журнал ВГПУ «Грани познания», 2011. № 2 (12). С. 1–9.

6. Захаров В. П., Богданова С. Ю. Корпусная лингвистика : учебник для студентов гуманитарных вузов. Иркутск: ИГЛУ, 2011. 161 с.

7. Плунгян В. А. Почему современная лингвистика должна быть лингвистикой корпусов [Электронный ресурс] // URL: polit.ru/articles/publichnye-lektsii/pochemu-sovremennaya-lingvistika-dolzha-byt-lingvistikoy-k-orpusov-2009-10-23/ (дата обращения 20.06.2024).

8. Герд А. С., Захаров В. П. Национальный корпус русского языка в свете проблем современной филологии // Корпусная лингвистика-2004 : сб. трудов междунар. конф., С.-Петербург, 11–14 окт. 2004 г. / С.-Петербург. гос. ун-т. редкол.: Л. Н. Беляева (отв. ред.) [и др.]. СПб., 2004. С. 122–131.

9. Sinclair J. Corpus, Concordance, Collocation. Oxford : Oxford Univ. Press, 1991. 200 p.

10. Ключихин В. В. Формирование коллокационной компетенции обучающихся на основе электронного лингвистического корпуса // Вестник Тамбовского университета. Сер. Гуманитарные науки, 2019. Т. 24. № 179. С. 69–80.

11. Сысоев П. В. Современные информационные и коммуникативные технологии: дидактические свойства и функции // Язык и культура, 2012. № 1 (17). С. 120–133.

12. Груздев Д. Ю. Электронный корпус текстов как эффективный инструмент переводчика : дис. ... канд. филол. наук : 10.02.19. М., 2013. 188 л.

13. Бусел Т. В. Электронный корпус текстов как альтернатива словарю // Тэарэтычныя і прыкладныя аспекты этналагічных даследаванняў : зб. навук. арт. ; пад навук. рэд.: Н.П. Мартысюк. Мінск: БНТУ, 2019. С. 407–414.

14. Владимов Н. В. Корпусный подход к решению переводческих проблем : автореф. дис. ... канд. филол. наук : 10.02.19. М., 2005. 20 с.