УДК 372.881.111.1
DOI: 10.17223/19996195/54/16

ИСПОЛЬЗОВАНИЕ ЯЗЫКОВЫХ КОРПУСОВ В ОБУЧЕНИИ ИНОСТРАННОМУ ЯЗЫКУ

О.Ю. Павлова

Аннотация. Рассматриваются некоторые возможности использования электронного языкового корпуса в преподавании иностранного языка на примере Корпуса современного американского варианта английского языка Марка Дэвиса (Corpus of Contemporary American English) как наиболее репрезентативного корпуса, находящегося в открытом доступе. Определены основные характеристики языкового корпуса и обоснованы его преимущества по сравнению со словарями и интернетом, рассматриваемым в качестве обширного корпуса текстов. Представлена классификация языковых корпусов по разным признакам. Особое внимание уделено учебным корпусам, состоящим из текстов, написанных не носителями языка, а изучающими данный язык как иностранный. Рассмотрена их роль в создании словарей, грамматик, справочных и учебных материалов на основе анализа типичных ошибок, допускаемых носителями того или иного языка в силу различий в системах родного и иностранного языков. Сформулированы основные научно-исследовательские и практические задачи, которые помогает решить языковой корпус, в том числе в переводоведении и практике перевода, а также в педагогической деятельности при обучении иностранному языку: лексикографические (компенсаторные и верификационные) задачи, призванные компенсировать недостаток информации в словарях; словообразовательные задачи, направленные на изучение деривационного потенциала языковых единиц; переводческие задачи, направленные на проверку переводческих гипотез о лексической сочетаемости, и некоторые другие. Приводятся разные типы разметки, обеспечивающие возможности использования языкового корпуса для решения поставленных задач. Представлена характеристика Корпуса современного американского варианта английского языка, его состава и структуры. Определено место данного корпуса в рамках разработанной классификации языковых корпусов. Сформулированы некоторые рекомендации по использованию Корпуса современного американского варианта английского языка в преподавании английского языка. В частности, предложен ряд примерных заданий, которые можно адаптировать под конкретные учебные или исследовательские цели, в том числе задания по поиску коллокаций (проверке лексической сочетаемости), поиску синонимов и проверке их сочетаемости с другими языковыми единицами, изучению деривационных возможностей слов, составлению тематических списков ключевых слов, поиску иллюстративного материала и работе с конкордансом.

Ключевые слова: языковой корпус; лингвистический корпус; корпус текстов; разметка; обучение иностранному языку; лексическая сочетаемость; коллокация

Введение

Новые информационные технологии прочно вошли в нашу жизнь, став необходимым инструментом в осуществлении практически всех направлений человеческой деятельности, в том числе в преподавании иностранного языка. Важность владения навыками использования информационных технологий подтверждается, в частности, включением одноименной дисциплины во все образовательные стандарты нового поколения.

Одним из таких инструментов является электронный языковой (лингвистический) корпус. Навыки работы с корпусными данными крайне необходимы современному преподавателю иностранного языка наряду с навыками использования электронных словарей, машинных переводчиков, интернет-ресурсов и других технических средств. Следует также отметить, что помимо расширения дидактических возможностей применение корпусных технологий также стимулирует интерес учащихся к учебному процессу и исследовательской деятельности, способствуя тем самым повышению их уровня владения иностранным языком.

В данной статье рассматриваются некоторые возможности использования языкового корпуса в преподавании иностранного языка на примере Корпуса современного американского варианта английского языка (Corpus of Contemporary American English, далее корпус COCA). Выбор корпуса COCA обусловлен тем, что на сегодняшний день – это крупнейший электронный корпус современного английского языка, находящийся в открытом доступе.

Методология исследования

При выполнении данной работы использовался общенаучный метод анализа письменных и электронных источников, а также применялись методы сравнения, классификации и систематизации полученной информации. На основе изученных источников автором сформулировано собственное определение языкового корпуса, составлена классификация языковых корпусов по разным признакам, описаны основные характеристики языкового корпуса, виды разметки и цели его использования. Далее в статье рассматриваются структура и состав корпуса COCA, анализируются его возможности и преимущества в педагогической деятельности, а также предлагаются примерные задания по использованию данного корпуса при обучении английскому языку.

Вопросы включения лингвистических корпусов в методику преподавания иностранного языка активно обсуждаются в современной лингводидактике в последнее десятилетие. Неоспоримой считается роль корпусов в формировании иноязычной коммуникативной компе-

тенции обучающихся. В частности, языковые корпусы предлагается использовать для изучения значения слов, полисемии, употребления близких по семантике слов, управления глаголов, грамматических времен, а также для составления лексико-грамматических тестов [1–5]. Актуальным является также создание корпусов предметной области, предполагающее отбор высоковероятных речевых образцов профессионального общения, что «оптимизирует формирование и совершенствование лексических навыков и умений и создает условия для внедрения рациональной методики обучения иностранному языку» [6. С. 216]. Использование параллельных корпусов, т.е. корпусов исходных текстов и их переводов на другой язык (языки), способствует формированию лексических навыков на основе перевода [7].

Помимо лексических навыков, корпус позволяет совершенствовать и грамматические навыки обучающихся, в том числе образовывать грамматические конструкции в зависимости от коммуникативной интенции и ситуации общения, различать грамматическое оформление письменных и устных текстов, устанавливать логические связи между элементами предложения [3]. Еще одно преимущество так называемого корпусного подхода к обучению иностранным языкам состоит в том, что самостоятельная работа обучающихся с языковым корпусом способствует развитию не только рецептивных, но и продуктивных навыков, т.е. навыков производства письменной и устной речи в соответствии с условиями и требованиями конкретной ситуации профессиональной деятельности [8, 9]. Наконец, работа с лингвистическим корпусом активизирует и исследовательскую составляющую учебного процесса, что повышает мотивацию обучающихся к изучению иностранного языка [10].

Исследование и результаты

Проанализировав представленные в научной литературе определения языкового корпуса [11–15], можно вывести следующее обобщенное определение: языковой корпус – это репрезентативная выборка текстов или их фрагментов, представляющая собой определенным образом унифицированный, структурированный и размеченный текстовый массив конечного объема, доступный в электронном формате и предназначенный для решения конкретных исследовательских задач.

Интересно отметить, что интернет, по своей сути, – тот же корпус, только в разы больше любого языкового корпуса. Тем не менее в состав языкового корпуса, в отличие от интернета, входят только аутентичные тексты, что гарантирует достоверность результатов поиска. Исключение составляют, например, учебные корпусы, представлен-

ные текстами, созданными не носителями языка, а иностранцами, изучающими данный язык (подробнее см. ниже).

Одной из главных характеристик языкового корпуса, также выгодно отличающих его от сети Интернет, является его репрезентативность (или сбалансированность), под которой понимается пропорциональное представление в корпусе текстов разных периодов и жанров, что в свою очередь обеспечивает необходимую полноту и типичность представленных в нем языковых данных.

Управление текстовыми данными в корпусе осуществляется с помощью корпусного менеджера (corpus manager) – «специализированной поисковой системы, включающей программные средства для поиска данных в корпусе, получения статистической информации и предоставления пользователю результатов в удобной форме» [11. С. 8]. В данной связи интересным представляется следующее наблюдение, сделанное британским лингвистом Сьюзен Ханстон: «…хотя в корпусе не содержится новой информации о языке, с помощью программного обеспечения, используемого для обработки данных, можно получить новый взгляд на то, что уже известно» (здесь и далее перевод мой. – *О.П.*) [16. С. 2–3].

Возможность использования языкового корпуса для решения разного рода исследовательских задач обеспечивается процедурой разметки (annotation, tagging), т.е. приписывания языковым единицам определенных характеристик (грамматических, семантических и др.) на этапе обработки текстовых данных при составлении корпуса. Выделяют лингвистическую и экстралингвистическую разметку. Лингвистическая разметка представлена следующими видами: морфологическая, синтаксическая, семантическая, анафорическая (референтные связи), просодическая (ударение и интонация) и дискурсная (паузы, повторы, оговорки) [11. С. 45–50]. Экстралингвистическая разметка включает в себя «внешнюю», «интеллектуальную» разметку (библиографические, типологические, тематические, социологические характеристики), «формальную» структурную разметку (текст, раздел, глава, часть, абзац, предложение) и технико-технологическую разметку (кодировка, даты обработки, исполнители, источник электронной версии) [11. С. 50].

Подробная размеченность языкового корпуса позволяет использовать его для решения широкого спектра научно-исследовательских и практических задач, в том числе в переводоведении и практике перевода, а также при обучении иностранному языку. Подводя итоги конференции «Национальный корпус русского языка и проблемы гуманитарного образования» (2007 г.), Т.В. Шмелёва составила своего рода «задачник» корпуса, обобщающий задачи, которые помогает решать языковой корпус. В частности, ею выделены следующие задачи: 1) лексикографические (например, поиск иллюстративного материала, анализ

изменений в семантике слов, создание новых словарей и т.д.), которые в свою очередь подразделяются на а) компенсаторные (восполнение дефицита данных в словарях) и б) верификационные (проверка данных, имеющихся в словарях); 2) вариантностные, предполагающие изучение имеющихся вариантов (например, грамматического рода – *штиблет/штиблета*, падежных флексий – *народа/народу*, морфемной структуры глаголов – *осмысливать/осмыслять* и т.д.); 3) словообразовательные (например, изучение продуктивности словообразовательных моделей, контекстуальных возможностей производных слов, деривационной вариантности и т.д.); 4) синтаксические (например, изучение вариантности глагольного управления – *тосковать о/тосковать по*) и 5) орфографические задачи, позволяющие фиксировать современные тенденции в области правописания (например, *бренд/брэнд*) [17. С. 26–32].

На наш взгляд, вариантностные задачи, входящие в вышеприведенный «задачник», представляют собой задачи несколько иного порядка по сравнению с остальными, поскольку они предполагают поиск и изучение представленных в корпусе вариантов языковых единиц на разных уровнях: лексическом, морфологическом, синтаксическом и т.д. Таким образом, их целесообразно рассматривать не как отдельный тип задач, а в качестве подзадач, сопровождающих решение конкретной лингвистической задачи. То же можно сказать и о лексикографических задачах (за исключением непосредственного составления словарей), поскольку любая работа с языковым корпусом так или иначе позволяет подтвердить, дополнить или скорректировать данные разных языковых уровней, представленные в лексикографических источниках.

Существование множества языковых корпусов позволяет классифицировать их по ряду признаков (табл. 1). Представленная классификация составлена на основе классификаций В.В. Рыкова [12] и В.П. Захарова и С.Ю. Богдановой [11. С. 21–22].

С точки зрения преподавания и изучения иностранного языка важно отметить роль языковых корпусов при составлении словарей, грамматик, справочных и учебных материалов. Как известно, все современные толковые словари английского языка, предназначенные для изучающих этот язык как иностранный (Longman Dictionary of Contemporary English, Cambridge Advanced Learner's Dictionary, Collins COBUILD English Language Dictionary и др.), основаны на корпусных данных, т.е. являются корпусно-ориентированными. Это означает, что в состав словарной статьи включаются те значения языковых единиц, те коллокации и идиоматические выражения, те грамматические конструкции, которые наиболее часто реализуются носителями языка в современной речи, т.е. соответствуют современному узусу. В качестве иллюстративного материала приводятся не искусственно созданные лексикографами примеры, не отрывки из классической художественной

литературы, а примеры из реальных современных текстов, которые придется продуцировать изучающему иностранный язык в повседневной жизни и профессиональной сфере. Кроме того, анализ статистических данных, полученных с помощью корпуса, позволяет выявлять соответствующие особенности языковых явлений в зависимости от формы (устной или письменной), регистра, стиля, жанра и т.д. Все это непременно повышает эффективность изучения иностранного языка.

Таблица 1

Классификация языковых корпусов

Признак	Типы корпусов
Тип языковых данных	Устные, письменные, мультимедийные, смешанные
Язык представления текстов	Одноязычные, двуязычные, многоязычные, смешанные
Языковая норма	Литературные, диалектные, смешанные
Дискурсивно-стилистическая принадлежность текстов	Литературные, фольклорные, разговорные, публицистические, научные (терминологические), смешанные
Жанрово-родовая принадлежность текстов	Прозаические, поэтические, драматургические, смешанные
Цель использования	Многоцелевые, специализированные
Способ доступа	Свободно доступные, коммерческие, закрытые, смешанные
Назначение	Исследовательские, иллюстративные, смешанные
Хронологический признак	Синхронические, диахронические
Способ существования корпуса	Динамические (пополняемые), статические
Индексация	Размеченные (аннотированные), неразмеченные
Характер разметки	Морфологические, синтаксические, семантические, просодические, смешанные
Объем текстов	Полнотекстовые, фрагментотекстовые, смешанные

Особого внимания в данном контексте заслуживают так называемые учебные корпусы (learner corpus), состоящие из текстов, написанных не носителями языка, а изучающими данный язык (сочинения, экзаменационные материалы и т.д.). Учебные корпусы используются при изучении процесса усвоения второго языка, а также в контрастивных исследованиях при сопоставлении речи носителей языка с речью изучающих данный язык [18. С. 103].

Одним из результатов таких исследований являются учебные материалы, в которых приводятся рекомендации по изучению языка, составленные на основе анализа типичных ошибок, допускаемых носителями того или иного языка в силу различий в системах этих языков. В качестве примера можно привести пособие для подготовки к ЕГЭ по английскому языку *Exam Success* [19], составленное на основе материалов Кембриджского учебного корпуса (Cambridge Learner Corpus).

В данном пособии есть рубрика *Get it right!*, где разъясняются конкретные трудности, с которыми могут столкнуться носители русского языка, и приводятся их типичные ошибки. Так, в частности, обращается внимание на то, что русскоязычные учащиеся часто образуют отрицательные прилагательные путем добавления приставки -un (например, *unuseful* вместо *useless*) [19. С. 124]. Очевидно, что использование таких пособий способствует повышению качества языковой подготовки, помогая учащимся лучше ориентироваться в современном английском языке и выражать свои мысли более узуально, не копируя привычные формы родного языка, что приводит к созданию текстов на так называемом *Russian English*.

Корпус COCA был создан Марком Дэвисом, профессором корпусной лингвистики из Университета Бригама Янга [20]. С декабря 2017 г. (последнее обновление) объем корпуса составляет свыше 560 млн слов (словоупотреблений или токенов) и хронологически охватывает период 1990–2017 гг. Корпус ежегодно пополняется на 20 млн слов. Структуру и состав корпуса можно представить в виде табл. 2.

Таблица 2

Структура и состав корпуса COCA

Название подкорпуса	Объем подкорпуса, млн слов (%)	Состав подкорпуса
Устная речь (Spoken)	118 (20,5)	Транскрипты теле- и радиопрограмм, например *All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Today Show (NBC), 60 Minutes (CBS), Hannity and Colmes (Fox) и др.*
Художественная литература (Fiction)	113 (19,7)	Рассказы и пьесы из литературных, детских и популярных журналов, первые главы первых изданий книг (с 1990 г. по наст. вр.), сценарии фильмов
Популярные журналы (Popular Magazines)	118 (20,5)	Почти 100 различных журналов (новостных, деловых, женских, спортивных, религиозных и др.), например *Time, Men's Health, Good Housekeeping, Cosmopolitan, Fortune, Christian Century, Sports Illustrated и др.*
Пресса (Newspapers)	114 (19,8)	10 американских газет, в том числе *USA Today, New York Times, Atlanta Journal Constitution, San Francisco Chronicle* и др.
Научные журналы (Academic Journals)	112 (19,5)	Около 100 рецензируемых журналов в области философии, психологии, религии, истории, технологии, образования и др.

Как видно из табл. 2, корпус COCA является репрезентативным и пропорционально сбалансированным по своему составу, что позволяет в том числе проводить достоверные компаративные исследования.

В соответствии с классификацией языковых корпусов, представленной в табл. 1, COCA является одноязычным, литературным по языковой норме, прозаическим, многоцелевым, свободно доступным, динамическим, размеченным корпусом, смешанным по типу языковых данных (устные и письменные тексты), дискурсивно-стилисти-ческой принадлежности текстов (литературные, разговорные, публицистические и научные тексты), назначению (материал для исследования или иллюстративный материал), характеру разметки (морфологическая, синтаксическая и семантическая разметка) и объему текстов (полные тексты и фрагменты текстов).

Многомиллионный размер корпуса, безусловно, является его преимуществом. Вместе с тем в профессиональном педагогическом сообществе обсуждается проблема недостатка специализированных тематических корпусов. В связи с этим Майкл Маккарти (британский лингвист, профессор прикладной лингвистики, автор и соавтор популярных пособий *English Collocations in Use* и *English Vocabulary in Use* издательства *Cambridge University Press*) отмечает необходимость привлечения учителей к созданию корпусов и корпусных материалов, важность диалога между учителями и разработчиками корпусов. В частности, он призывает учителей «формулировать свои пожелания к корпусам и ресурсам, перестать быть немым, пассивным адресатом основанных на корпусе образовательных ресурсов. Пока что такой диалог совершенно не развит; но уже сейчас можно отметить, что преподаватели… говорят о необходимости создания специализированных, тематических корпусов. Обычные диверсифицированные корпуса оказываются недостаточными для педагогической деятельности: даже крупный, многомиллионный корпус содержит недостаточно текстов по сельскохозяйственной тематике для того, чтобы студент смог работать со специфической для этой области лексикой» [21. С. 41].

В корпусе COCA проблема создания специализированных тематических корпусов отчасти решается за счет возможности поиска какого-либо слова или словосочетания не во всем корпусе, а в текстах конкретного подкорпуса (например, научного) или даже в текстах из конкретного тематического источника (например, в научных текстах по медицине, юриспруденции, истории, образованию, религии и другим темам). Более того, есть возможность сравнить употребление языковой единицы в текстах разных временных периодов, жанров или тематики.

Ниже представлены некоторые рекомендации по использованию корпуса COCA в преподавании английского языка. В частности, предложены примерные типы заданий, которые можно адаптировать под те

или иные цели – учебные, исследовательские или же просто иллюстративные. Разумеется, что языковой корпус можно и нужно использовать и для решения переводческих задач, в частности для проверки переводческих гипотез о лексической или лексико-грамматической сочетаемости.

1. *Проверка лексической сочетаемости (поиск коллокаций).*

Под коллокацией в широком смысле понимается «комбинация двух или более слов, имеющих тенденцию к совместной встречаемости» [22. С. 137].

Владение коллокациями считается настолько важным компонентом языковой компетенции, что в современной дидактической литературе выделяется даже понятие «коллокационная компетенция», под которой понимается «умение правильно и точно сочетать слова с соблюдением норм иностранного языка» [23. С. 77]. Языковой корпус представляется незаменимым средством формирования коллокационной компетенции, поскольку позволяет в считанные секунды получить всеобъемлющую информацию о сочетаемости того или иного слова и дает возможность просмотреть множество контекстов употребления искомых коллокаций.

В этой связи следует упомянуть о существовании словарей коллокаций [24]. Несмотря на то что они дают достаточно полное представление о лексико-грамматической сочетаемости того или иного слова, в них, как правило, не указывается частотность тех или иных словосочетаний. Вместе с тем языковой корпус, в частности корпус COCA, позволяет подобрать наиболее частотную и, следовательно, типичную коллокацию в текстах определенного жанра, стиля и т.д. Так, задав поиск коллокаций по запросу «adj. + problem» (табл. 3), мы можем увидеть, что самым частотным прилагательным в данной модели словосочетания является *big* (4 200 употреблений в корпусе), и встречается оно преимущественно в устной речи (2129), достаточно часто – в прессе (855) и журналах (763), реже всего – в научных текстах (196). Для художественной литературы данная коллокация характерна в наименьшей степени. Таким образом, если нам необходимо подобрать словосочетание для научного текста, то, опираясь на данные частотности, мы скорее предпочтем варианты *serious problem* или *major problem*, для публицистического текста – *big problem* или *real problem* и т.д. Заметим, что подобной информации нет ни в одном словаре; таким образом, корпус в данном случае выполняет компенсаторную функцию, восполняя недостаток данных в лексикографических источниках.

Учитывая вышеизложенное, можно порекомендовать следующие примерные задания на поиск коллокаций: а) изучите сочетаемость прилагательного *economic* с существительными; б) сравните употребление глаголов *take* и *make* с существительными (функция Compare); в) определите, какие прилагательные чаще всего употребляются с существи-

тельными *man* и *woman* в художественных и публицистических текстах, и т.д.

Таблица 3

Частотность коллокации «adj. + problem» в корпусе COCA

Прилагательное	Всего	Spoken	Fiction	Magazine	Press	Academic
Big	4 200	**2 129**	227	**763**	**885**	196
Real	2 374	**1 206**	181	**373**	**396**	218
Serious	1 619	**599**	81	291	256	**392**
Major	1 282	**382**	35	245	245	**375**
Only	1 203	**333**	236	**311**	245	78
Main	593	**154**	45	**131**	**137**	**126**
Huge	572	**363**	15	72	94	28
Fundamental	452	**173**	4	72	63	**140**
Common	451	67	9	**172**	57	**146**

2. *Поиск синонимов.*

Как и в случае со словарями коллокаций, словари синонимов не дают информации о частотности лексических единиц, что легко компенсируется языковым корпусом. Помимо построения синонимических рядов и подбора синонимичных единиц в зависимости от жанрово-стилистических особенностей текста, языковой корпус можно использовать для изучения сочетаемости синонимичных слов с другими словами. В связи с этим можно предложить учащимся, например, следующие задания: а) составьте список синонимов прилагательного *global* (в корпусе COCA синоним обозначается знаком равенства, т.е. запрос будет выглядеть так: *=global*); б) сравните употребление синонимичных прилагательных *fast* и *rapid* с существительными; в) изучите сочетаемость синонимичных глаголов *build* и *construct* с существительными, обозначающими объект строительства; г) сравните употребление существительных со значением снижения/падения в устной речи и научном подкорпусе и т.д. Так, результаты поиска по последнему запросу показали, что в устной речи чаще встречаются существительные *cut, drop* и *fall*, в то время как в научных текстах преобладают существительные *decrease*, *diminution*, *decline* и *reduction.*

3. Изучение деривационных возможностей слов.

Как было сказано выше, языковой корпус помогает решать и словообразовательные задачи, в том числе определять продуктивность словообразовательных моделей, изучать деривационные возможности языковых единиц, выявлять словообразовательные неологизмы и т.д. В связи с этим учащимся можно предложить следующие задания: а) изучите деривационный потенциал корня *clean* (пропущенная часть слова обозначается звездочкой: *clean*); б) составьте список всех прилагательных с суффиксом *-less*; в) какие слова могут занимать позицию

первого компонента в составном прилагательном “noun/adv.-friendly” (по аналогии с *user-friendly/environmentally friendly*) и т.д.

4. Составление тематических списков ключевых слов.

Оформив соответствующий запрос в корпусе COCA, учащиеся могут в считаные секунды получить списки наиболее частотных слов (при необходимости заданной части речи) в текстах определенного жанра, стиля или тематики, а также конкретного года или периода времени. Данная возможность делает языковой корпус незаменимым помощником при изучении лексики. Примерами заданий могут быть следующие: а) составьте списки наиболее частотных прилагательных, употребляемых в религиозных и развлекательных журналах (в сравнении); б) составьте список существительных, наиболее часто употребляемых в научных журналах по медицине и т.д. Следует также отметить, что у пользователя есть возможность сохранить или скачать составленные списки слов для дальнейшего использования.

5. Поиск иллюстративного материала.

Одно из главных преимуществ языкового корпуса состоит в том, что языковые данные представлены там в своем естественном контексте, что дает возможность изучения как непосредственного лексического окружения языковых единиц, так и их функционирования в более широком контексте на уровне фрагмента текста. Контексты употребления искомых языковых единиц можно использовать в качестве иллюстративного материала для объяснения изучаемых языковых явлений или для получения более полного представления о слове или словосочетании по сравнению с тем, что способны дать словари. Важно отметить, что все фрагменты текстов, предлагаемые корпусом COCA в качестве иллюстративного контекста, сопровождаются подробными выходными данными с указанием автора, названия источника, года публикации, издательства, страниц и названия соответствующего подкорпуса.

Для изучения лексического окружения искомого слова в корпусе COCA можно запросить представление результатов поиска в виде конкорданса (KWIC, key word in context). Конкорданс – это перечень (как правило, алфавитный) всех случаев употребления искомого слова в языковом корпусе в виде контекстов, в которых представлены цепочки из нескольких слов, стоящих слева и справа от искомого слова [18. С. 42–43]. Фрагмент конкорданса, построенного корпусным менеджером COCA на существительное *research*, представлен на рис. 1.

Как видно на рис. 1, искомое слово располагается по центру, а слева и справа от него разным цветом выделены различные части речи. Пролистав такой конкорданс, можно получить всестороннее представление о реальном употреблении данного слова в современном американском варианте английского языка. Следует также отметить, что такие важные характеристики корпуса, как сбалансированность и репре-

зентативность, обеспечивают пропорциональное представление результатов поиска в выборке и, соответственно, типичность описанных данных.

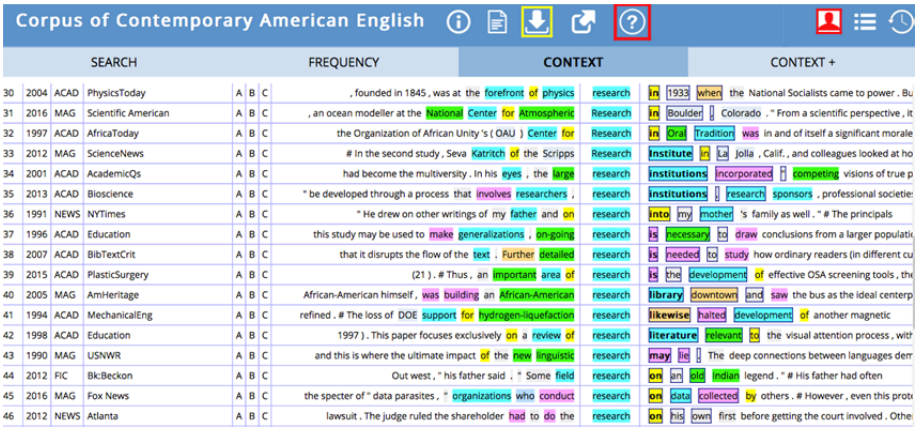

Рис. 1. Конкорданс существительного *research* в корпусе COCA

Заключение

В данной статье продемонстрирован ряд возможностей использования электронного языкового корпуса (на примере корпуса COCA) в педагогической деятельности. Преимущества размеченного языкового корпуса по сравнению с интернетом заключаются в аутентичном характере представленных в нем текстов, а также в возможностях программного обеспечения, позволяющего пользователю формулировать свой запрос в соответствии с конкретными целями и задачами. Кроме того, с помощью языкового корпуса можно в считанные секунды получить данные о частотности и, соответственно, узуальности той или иной языковой единицы или грамматической конструкции в текстах определенного жанра и временного периода. Эти данные существенно дополняют информацию, представленную в электронных словарях, что делает языковой корпус незаменимым помощником в изучении иностранного языка.

Языковой корпус позволяет лингвистам, переводчикам и преподавателям решать целый ряд лексикографических (компенсаторных и верификационных), словообразовательных, переводческих, дидактических и научно-исследовательских задач. Учебные корпуса активно используются в создании современных словарей, грамматик, справочных и учебных материалов, позволяя анализировать типичные ошибки, допускаемые носителями того или иного языка и, соответственно, помогая им избегать этих ошибок.

В работе сформулированы некоторые рекомендации по использованию корпуса COCA в преподавании английского языка. В частности, предложен ряд примерных заданий, которые можно адаптировать под конкретные учебные или исследовательские цели, а именно задания по поиску коллокаций или проверке лексической сочетаемости, по поиску синонимов и проверке их сочетаемости с другими языковыми единицами, по изучению деривационных возможностей слов, составлению тематических списков ключевых слов, поиску иллюстративного материала и работе с конкордансом.

Литература

1. ***Дерябина И.В.*** Методика обучения учащихся управлению английских глаголов на основе британского национального корпуса (профильный уровень) : автореф. дис. ... канд. пед. наук. М., 2013.
2. ***Дерябина И.В.*** Характеристики педагогической работы с английским национальным корпусом // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2012. № 9. С. 156–160.
3. ***Рязанова Е.А.*** Методика формирования грамматических навыков речи студентов на основе лингвистического корпуса : автореф. дис. ... канд. пед. наук. М., 2012.
4. ***Сысоев П.В.*** Лингвистический корпус в методике обучения иностранным языкам // Язык и культура. 2010. № 1 (9). С. 99–111.
5. ***Чернякова Т.А.*** Использование лингвистического корпуса в обучении иностранному языку // Язык и культура. 2011. № 4 (16). С. 127–132.
6. ***Горина О.Г.*** Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке : дис. ... канд. пед. наук. СПб., 2014. 308 с.
7. ***Кокорева А.А.*** Корпус параллельных текстов в обучении иностранному языку // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2013. № 2 (118). С. 57–62.
8. ***Толстова Т.В.*** Корпусная лингвистика в обучении английскому языку делового общения // Вопросы прикладной лингвистики : сб. науч. трудов. М. : Рос. ун-т дружбы народов, 2009. С. 63–67.
9. ***Базарова Б.Б.*** Возможности корпусной лингвистики в языковом обучении // Евразийская парадигма России: ценности, идеи, практика : материалы междунар. науч. конф., посвящ. 20-летию Бурятского государственного университета. Улан-Удэ : Изд-во Бурят. гос. ун-та, 2015. С. 90–92.
10. ***Мордовин А.Ю.*** Лингвистическое исследование как неотъемлемая часть применения национальных корпусов текстов в обучении иностранным языкам // Гуманитарный вектор. 2012. № 4 (32). С. 274–279.
11. ***Захаров В.П., Богданова С.Ю.*** Корпусная лингвистика. Иркутск : ИГЛУ, 2011. 161 с.
12. ***Рыков В.В.*** Корпусная лингвистика: курс лекций. URL: http://rykov-cl.narod.ru/c.html (дата обращения: 22.01.2020).
13. ***Плунгян В.А.*** Почему современная лингвистика должна быть лингвистикой корпусов // Публичная лекция. 2009. URL: www.polit.ru/lectures/2009/10/23/corpus.html (дата обращения: 22.01.2020).
14. ***McEnery T., Wilson A.*** Corpus Linguistics. Edinburgh University Press, 2001.
15. ***Meyer Ch.*** English Corpus Linguistics. An Introduction. Cambridge University Press, 2004. 168 p.

16. ***Hunston S.*** Corpora in Applied Linguistics. Cambridge : Cambridge University Press, 2002.
17. ***Шмелева Т.В.*** Задачник корпуса // Национальный корпус русского языка и проблемы гуманитарного образования / под ред. Н.Р. Добрушиной. М., 2007. С. 25–34.
18. ***Baker P., Hardie A., McEnery T.A.*** Glossary of Corpus Linguistics. Edinburgh University Press, 2006. 187 p.
19. ***Rimmer W., Vinogradova O., Kozhevnikova L.*** Exam Success. Подготовка к ЕГЭ по английскому языку. Cambridge University Press, 2013. 296 p.
20. ***Davies M.*** The Corpus of Contemporary American English. URL: https://www.english-corpora.org/coca/ (дата обращения: 22.01.2020).
21. ***McCarthy M.*** Accessing and interpreting corpus information in the teacher education context // Language Teaching. Vol. 41. Cambridge University Press, 2008. P. 563–574. URL: https://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4-child/texts/mccar-thy_teachers.pdf
22. ***Захаров В.П., Хохлова М.В.*** Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии. 2010. № 9 (16). С. 137–143. URL: http://www.dialog-21.ru/digests/dialog2010/materials/html/22.htm (дата обращения: 22.01.2020).
23. ***Клочихин В.В.*** Формирование коллокационной компетенции обучающихся на основе электронного лингвистического корпуса // Вестник Тамбовского университета. Серия: Гуманитарные науки. 2019. Т. 24, № 179. С. 68–80.
24. ***Oxford*** Collocations Dictionary for students of English. Oxford University Press, 2010. 897 p.

Сведения об авторе:
Павлова Оксана Юрьевна – кандидат филологических наук, доцент, Кемеровский государственный университет (Кемерово, Россия). E-mail: oxanashishigina@rambler.ru

Поступила в редакцию 15 апреля 2021 г.

Linguistic corpora in foreign language teaching

Pavlova O.Yu., Ph.D. (Philology), Associate Professor, Kemerovo State University (Kemerovo, Russia). E-mail: oxanashishigina@rambler.ru

DOI: 10.17223/19996195/54/16

Abstract. This article discusses some possible uses of electronic linguistic corpora in foreign language teaching. It looks at the Corpus of Contemporary American English created by Mark Davies as the largest open access corpus of English available today. The paper defines what a linguistic corpus is and describes its main characteristics. It focuses on the advantages of using corpora over dictionaries, including dictionaries of collocations and synonyms, and the Internet considered as an enormous corpus of texts. The article offers a classification of linguistic corpora according to various criteria. Particular attention is paid to learner corpora, which consist of texts written by learners of a language rather than its native speakers, and their role in studies of second language acquisition. Learner corpora help to create dictionaries, grammars, textbooks, and reference materials based on the analysis of common errors made by language learners. They provide special recommendations or guidelines to particular language speakers to prevent them from making typical mistakes caused by differences in the systems of their native language and the one they are learning. Further, the paper lists possible research objectives and practical tasks that linguistic corpora can help us fulfil in translation studies and teaching a foreign language. These include lexicographic objectives meant to verify the data given in the dictionaries and compensate for the lack of valuable information on language use and frequency of occurrence there. Linguistic corpora

are also useful in studying word-formation patterns and the derivational potential of linguistic units. In addition, they are extremely instrumental in helping translators check their hypotheses and use the most natural collocations and grammatical structures. The paper explains how linguistic corpora are annotated or tagged to enable them to help linguists, translators, teachers, and language learners use them for their purposes. Using the Corpus of Contemporary American English as an example, we describe its composition and structure and define its place in the classification of linguistic corpora that we developed. Finally, we offer some recommendations and sample tasks that can be given to language learners to help them answer their specific questions. These include searching for collocates (words that occur nearby a given word), building lists of synonyms and checking their compatibility with other words, studying the derivational potential of morphemes, creating lists of keywords on a certain topic, studying new words in their natural contexts through concordances, etc.
Keywords: linguistic corpus; learner corpus; annotation; foreign language teaching; collocation

References

1. Deryabina I.V. (2013) Metodika obucheniya uchashchikhsya upravleniyu angliyskikh glagolov na osnove britanskogo natsional'nogo korpusa [Methods of teaching English verb patterns to students based on the British National Corpus]. Abstract of Pedagogics cand. diss. M.
2. Deryabina I.V. (2012) Kharakteristiki pedagogicheskoy raboty s angliyskim natsional'nym korpusom [Characteristics of pedagogical work with the English National Corpus] // Vestnik Tambovskogo universiteta. Seriya 2: Gumanitarnyye nauki, 9. pp. 156–160.
3. Ryazanova E.A. (2012) Metodika formirovaniya grammaticheskikh navykov rechi studentov na osnove lingvisticheskogo korpusa [Methodology for the formation of students' grammar skills based on the linguistic corpus]. Abstract of Pedagogics cand. diss. M.
4. Sysoev P.V. (2010) Lingvisticheskiy korpus v metodike obucheniya inostrannym yazykam [Linguistic corpus in the methodology of teaching foreign languages] // Yazyk i kul'tura – Language and Culture, 1 (9). pp. 99–111.
5. Chernyakova T.A. (2011) Ispol'zovaniye lingvisticheskogo korpusa v obuchenii inostrannomu yazyku [The use of linguistic corpus in teaching a foreign language] // Yazyk i kul'tura – Language and Culture, 4 (16). pp. 127–132.
6. Gorina O.G. (2014) Ispol'zovaniye tekhnologiy korpusnoy lingvistiki dlya razvitiya leksicheskikh navykov studentov-regionovedov v professional'no-oriyentirovannom obshchenii na angliyskom yazyke [Using corpus linguistics technologies to develop vocabulary skills of the students majoring at Foreign Studies in profession-oriented communication in English]. Pedagogics cand. diss. SPb. 318 p.
7. Kokoreva A.A. (2013) Korpus parallel'nykh tekstov v obuchenii inostrannomu yazyku [A corpus of parallel texts in teaching a foreign language] // Vestnik TGU, 2 (118). pp. 57–62.
8. Tolstova T.V. (2009) Korpusnaya lingvistika v obuchenii angliyskomu yazyku delovogo obshcheniya [Corpus linguistics in teaching Business English] // Voprosy prikladnoy lingvistiki. M.: RUDN University. pp. 63–67.
9. Bazarova B.B. (2015) Vozmozhnosti korpusnoy lingvistiki v yazykovom obuchenii [Opportunities of corpus linguistics in language teaching] // Evraziyskaya paradigma Rossii: tsennosti, idei, praktika. Ulan-Ude: Buryat State University. pp. 90–92.
10. Mordovin A.Yu. (2012) Lingvisticheskoye issledovaniye kak neot'yemlemaya chast' primeneniya natsional'nykh korpusov tekstov v obuchenii inostrannym yazykam [Linguistic research as an integral part of using national corpora in teaching foreign languages] // Gumanitarnyy vector, 4 (32). pp. 274-279.

11. Zakharov V.P., Bogdanova S.Yu. (2011) Korpusnaya lingvistika [Corpus linguistics]. Irkutsk: IGLU.
12. Rykov V.V. Korpusnaya lingvistika: kurs lektsiy [Corpus linguistics: a course of lectures]. URL: http://rykov-cl.narod.ru/c.html (Accessed: 22.01.2020).
13. Plungyan V.A. (2009) Pochemu sovremennaya lingvistika dolzhna byt' lingvistikoy korpusov: publichnaya lektsiya [Why should modern linguistics be a corpus linguistics: public lecture]. URL: www.polit.ru/lectures/2009/10/23/corpus.html (Accessed: 22.01.2020).
14. McEnery T., Wilson A. (2001) Corpus Linguistics. Edinburgh University Press.
15. Meyer Ch. (2004) English Corpus Linguistics. An Introduction. Cambridge University Press. 168 p.
16. Hunston S. (2002) Corpora in Applied Linguistics. Cambridge: Cambridge University Press.
17. Shmeleva T.V. (2007) Zadachnik korpusa [Ehe set of problems for corpus] // Natsional'nyy korpus russkogo yazyka i problemy gumanitarnogo obrazovaniya [The Russian National Corpus and the problems of humanitarian education]. Ed. by N.R. Dobrushina. M. pp. 25–34.
18. Baker P., Hardie A., McEnery T. (2006) A glossary of Corpus Linguistics. Edinburgh University Press. 187 p.
19. Rimmer W., Vinogradova O., Kozhevnikova L. (2013) Exam Success. Cambridge University Press. 296 p.
20. Davies M. The Corpus of Contemporary American English. URL: https://www.english-corpora.org/coca/ (Accessed: 22.01.2020)
21. McCarthy M. (2008) Accessing and interpreting corpus information in the teacher education context // Language Teaching, 41. Cambridge University Press. pp. 563–574 URL: https://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4-child/texts/mccarthy_teachers.pdf
22. Zakharov V.P., Khokhlova M.V. (2010) Analiz effektivnosti statisticheskikh metodov vyyavleniya kollokatsiy v tekstakh na russkom yazyke [Analysis of the effectiveness of statistical methods for identifying collocations in Russian texts] // Komp'yuternaya lingvistika i intellektual'nyye tekhnologii, 9 (16). pp. 137–143. URL: http://www.dialog-21.ru/digests/dialog2010/materials/html/22.htm (Accessed: 22.01.2020).
23. Klochikhin V.V. (2019) Formirovaniye kollokatsionnoy kompetentsii obuchayushchikhsya na osnove elektronnogo lingvisticheskogo korpusa [Development of students' collocation competence on the basis of electronic linguistic corpora] // Vestnik Tambovskogo universiteta: Gumanitarnyye nauki. 24 (179). pp. 68–80.
24. Oxford Collocations Dictionary for students of English (2010). Oxford University Press. 897 p.

Received 15 April 2021