

ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ПРОЦЕССЕ ИЗУЧЕНИЯ АНГЛИЙСКОГО ЯЗЫКА НА МАТЕРИАЛЕ КОРПУСНОЙ ЛИНГВИСТИКИ

USE OF INFORMATIONAL TECHNOLOGIES IN THE COURSE OF STUDYING OF ENGLISH ON THE MATERIAL OF CORPUS LINGUISTICS

© 2015 Гаджиева З. К.

Дагестанский государственный педагогический университет

© 2015 Gadzhiyeva Z. K.

Dagestan State Pedagogical University

Резюме. В данной статье раскрывается сущность использования корпусного подхода в процессе изучения английского языка на материале Britain National Corpus (BNC). Авторами изучены некоторые особенности применения корпусной лингвистики, в частности рассмотрены примеры использования текстовых корпусов того или иного диалекта английского языка.

Abstract. The article deals with the analysis of the essence of using the corpus approach in the process of studying English on the Britain national corpus (BNC) material. The author studied some features of applying the corpus linguistics, in particular examples of using the text cases of different English dialects.

Rezjume. V dannoj stat'e raskryvaetsya sushchnost' ispol'zovaniya korpusnogo podhoda v processe izucheniya anglijskogo yazyka na materiale Britain national corpus (BNC). Avtorami izucheny nekotorye osobennosti primeneniya korpusnoj lingvistiki, v chastnosti rassmotreny primery ispol'zovaniya tekstovyh korpusov togo ili inogo dialekta anglijskogo yazyka.

Ключевые слова: корпус, корпусная лингвистика, корпусный анализ, Британский национальный корпус.

Keywords: corpus, corpus linguistics, corpus analysis, British national corpus.

Klyuchevye slova: korpus, korpusnaya lingvistika, korpusnyj analiz, Britanskij nacional'nyj korpus.

В последнее время лингвисты все чаще используют информационные технологии, в частности методы корпусного анализа в решении своих когнитивных задач. Прогресс в области информационных и коммуникационных технологий влечет за собой прогресс в создании и совершенствовании программ автоматической обработки текста, т.е. корпусной лингвистики.

В настоящее время существует множество определений корпусной лингвистики.

Например, определение, приведенное в учебнике В. П. Захарова: **корпусная лингвистика** – раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов (корпусов текстов) с применением компьютерных технологий. Под **лингвистическим**, или **языковым, корпусом текстов** понимается унифицированный, структурированный, размеченный, филологически компетентный массив

языковых данных, предназначенный для решения конкретных лингвистических задач [1]. Т. Мак Энери и Э. Вилсон дают следующее определение: корпус – это собрание языковых фрагментов, отобранных в соответствии с четкими языковыми критериями для использования в качестве модели языка [5]. В. В. Рыков определяет корпус текстов как некоторое собрание текстов, в основе которого лежит логический замысел, логическая идея, объединяющая эти тексты и воплощенная в правилах организации текстов в корпус, алгоритме и программе анализа корпуса текстов, сопряженной с этим идеологии и методологии [3].

Таким образом, следует сказать, что корпусная лингвистика предоставляет материал для проведения различных исследований языка и его вариантов, и, тем самым, определяет основной метод анализа текстов на базе корпусов (Corpus-Based Approach) с применением информационных технологий со специализированными программами по обработке текстов на основе корпусного анализа. По мнению Е. П. Сосниной, корпусный подход, или метод лингвистического исследования, основанный на корпусах текстов, ориентирован на прикладное изучение языка, его функционирование в реальных средах и текстах, что важно для преподавания английского языка [4]. Корпусы текстов можно представить как коллекции текстов, специально отобранных, размеченных по тем или иным лингвистическим параметрам и обеспеченных системой поиска.

Необходимо также отметить, что основной целью использования информационных технологий, в частности программ, обеспечивающих работу с корпусами языков, в процессе изучения английского языка является создание и расширение национальных корпусов того или иного языка, представительных по отношению ко всему языку, которые могут служить для исследования самых разнообразных явлений этого языка.

Как объясняет А. В. Плунгян, термин «национальный корпус» восходит к английскому языку, так как первым был Британский национальный корпус, созданный в начале 90-х годов, который стал

эталонным корпусом в 90-е годы. Поэтому национальным корпусом стали называть всякий большой корпус данного языка [2].

На сегодняшний день существует большое многообразие языковых корпусов. Различают: The International Corpus of English (ICE) – Международный корпус английского языка, The British National Corpus – Британский национальный корпус, Synchronic corpora – синхронические корпуса, Historical corpora – исторические корпуса, The Freiburg English Dialect Corpus – Фрайбургский корпус английского диалекта, Michigan Corpus of Academic Spoken English (MICASE) – Мичиганский корпус академического разговорного английского языка.

Одним из наиболее полных проектов корпусов несомненно является The British National Corpus (BNC), включающий около 100 млн. слов. 90% единиц корпуса – из письменных источников, отобранных по трем основным критериям: область, время и формат. Под «областью» при этом понимается предметная область текста (например, «Финансы», «Прикладные науки», «Искусство» и другие); параметр «время» отражает дату/период публикации (1960-1974, 1975-1993 или «дата неизвестна»), а «формат» – вид публикации (например, книга, периодика, неопубликованные манускрипты). Меньшую часть (10%) корпуса составляет устный подкорпус, при отборе материала в который составители руководствовались двумя дополняющими друг друга параметрами – «демографическим» (для отбора спонтанных разговоров) и «контекстуальным» (для отбора записей встреч и передач). С одной стороны, для наиболее полного представления языка необходимо учитывать формат и контекст устной речи, поэтому половину устного подкорпуса BNC составляют неформальные разговоры, деловые встречи, лекции, радио передачи и др. Все источники можно разделить на четыре категории – образовательная/информативная (лекции, новостные и образовательные передачи – 20,56% источников), деловая (презентации товаров, встречи профсоюзов, собеседования – 21,47%), публичные речи (речи политиков, заседания парламента, проповеди – 21,86%) и

развлекательные (спортивные комментарии, заседания клубов, звонки на радио – 23,71%). Таким образом, можно отметить, что указанные четыре категории представлены практически равнозначно (12,38% источников остались 88 неклассифицированными). Вторая половина устной части корпуса – спонтанные разговоры, записанные 124 добровольцами в 38 городах Великобритании. Для демографически достоверной репрезентации речи британского общества в корпусе, в проекте участвовало одинаковое количество мужчин и женщин, практически равномерно представляющих различные возрастные группы и социальные классы. Все входные единицы корпуса BNC аннотированы, что позволяет производить поиск по частям речи, источникам и их характеристикам.

Для осуществления корпусного анализа на материале BNC пользователь нуждается в определенном интерфейсе. Самым распространенным вариантом является программа *XAIRA*, наряду с которым существуют и *BYU-BNC*, *BNC-Web* и *Intellitext*. Для решения лингвистических задач пользователю BNC прежде всего необходимо зарегистрироваться в окне интерфейса программы. После регистрации в системе, например, в *BNC-Web*, открывается главное меню. Здесь пользователь может задавать искомое слово или словосочетание в окне поиска. Можно также выбрать определенное количество результатов по заданному запросу и компьютер выводит результаты именно в конкретно заданном количестве. На следующем этапе работы после ввода слова пользователь получает статистические результаты по заданному запросу. Кнопка *ShowKWICVIEW* в окне

позволяет построить исследуемое слово в центре поля. В левой части таблицы синим шрифтом указаны данные об авторе текста: имя, пол, дата рождения (или примерный возраст); название текста; время и место создания текста (может указываться точно или приблизительно); объем текста.

Также можно осложнять условия запроса, задавая запрос не только по одному слову, но и по целым словосочетаниям, разным грамматическим конструкциям.

На базе корпусов компьютер обеспечивает возможность составить списки активной лексики студентов, частотные списки терминов для использования в профессиональных курсах обучения английскому языку. Благодаря тому, что BNC создается как открытый и пополняющийся корпус, то из него можно извлечь еще одну пользу: корпусные данные позволяют зафиксировать некоторые исторические вехи в развитии того или иного значения или коннотации у слов и словосочетаний.

Использование информационных технологий в процессе изучения английского языка дает эффективные результаты, так как обучающийся может легко найти те или иные корпуса языка (слова, словосочетания, афоризмы и др. языковые единицы) посредством специализированных программ по обработке тестов, встроенных в компьютер. Пользователь может получить данные о частоте встречаемости слов в том или ином контексте разного периода времени, получить данные о совместной встречаемости лексических единиц и т.д. Получив статистические данные, можно, например, выявить прототипические значения, выстроить их иерархию.

Литература

1. Захаров В. П. Корпусная лингвистика. СПб., 2005.
2. Плунгян В. А., Сичинава Д. В. Национальный корпус русского языка: опыт создания корпусов текстов современного русского языка. СПб., 2004.
3. Рыков В. В. Корпусная лингвистика (научно-аналитический обзор) // РЖ: Социальные и гуманитарные науки: Зарубежная литература. М., 1996. № 4.
4. Соснина Е. П. Введение в прикладную лингвистику. Ульяновск, 2000.
5. McEnery T. and Wilson A. *Corpus Linguistics*. (2nd Ed.) 2001. Edinburgh: Edinburgh University Press.
6. <http://www.titania.bham.ac.uk/docs/freq.gif>.
7. <http://www.dialog-21.ru/Archive/2003/Dobrovolskij.htm>.
8. <http://www.natcorp.ox.ac.uk/getting/sampler.html>

References

1. Zakharov V. P. *Corpus linguistics*. SPb., 2005.
2. Plungyan V. A., Sichinava D. V. *National corpus of Russian: experience of creation of corpuses of modern Russian*. SPb., 2004.
3. Rykov V. V. *Corpus linguistics (scientific state-of-the-art review)* // RJ: Social and humanities: Foreign literature. M, 1996. # 4.
4. Sosnina E. P. *Introduction to applied linguistics*. Ulyanovsk, 2000.
5. McEnery T. and Wilson A. *Corpus Linguistics*. (2nd Ed.) 2001. Edinburgh: Edinburgh University Press.
6. <http://www.titania.bham.ac.uk/docs/freq.gif>.

7. <http://www.dialog-21.ru/Archive/2003/Dobrovolskij.htm>
getting/sampler.html

8. [http:// www.natcorp.ox.ac.uk/](http://www.natcorp.ox.ac.uk/)

Literatura

1. Zaharov V. P. Korpusnaja lingvistika. SPb., 2005. 2. Plungjan V. A., Sichinava D. V. Nacional'nyj korpus russkogo jazyka: opyt sozdanija korpusov tekstov sovremennogo russkogo jazyka. SPb., 2004. 3. Rykov V. Korpusnaja lingvistika (nauchno-analiticheskij obzor) // RZh: Social'nye i gumanitarnye nauki: Zarubezhnaja literatura. M., 1996. № 4. 4. Sosnina E. P. Vvedenie v prikladnuju lingvistiku. Ul'janovsk, 2000. 5. McEnery T. and Wilson A. Corpus Linguistics. (2nd Ed.) 2001. Edinburgh: Edinburgh University Press. 6. <http://www.titania.bham.ac.uk/docs/freq.gif>. 7. [http://www.dialog-21.ru/Archive/2003 /Dobrovolskij.htm](http://www.dialog-21.ru/Archive/2003/Dobrovolskij.htm). 8. <http://www.natcorp.ox.ac.uk/getting/sampler.html>

Статья поступила в редакцию 06.07.2015 г.