
ТЕОРИЯ И МЕТОДИКА ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ THEORY AND METHODOLOGY OF PROFESSIONAL EDUCATION

Научная статья

УДК 378.14:811

DOI: 10.35854/2541-8106-2022-2-108-115

Совершенствование лексической компетенции студентов на основе работы с языковыми корпусами материала

Юлия Григорьевна Давыдова

Санкт-Петербургский университет технологий управления и экономики,
Санкт-Петербург, Россия, alexyulin@mail.ru

Аннотация. В статье рассмотрены возможности совершенствования лексической компетенции студентов на основе современных корпусов языкового материала. Автор уточняет значение понятия «корпус» в научных и прикладных исследованиях, обобщает взгляды ученых относительно ценности корпусных данных в лингвистических и лингводидактических целях, анализирует лингводидактический потенциал трех современных англоязычных корпусов. В статье приведены разнообразные задания, позволяющие студентам усовершенствовать знания в области семантики, сочетаемости, прагматики лексических единиц.

Ключевые слова: корпус, лингводидактический потенциал корпусной лингвистики, проблемный подход к обучению, лексическая компетенция, комплекс упражнений

Для цитирования: Давыдова Ю. Г. Совершенствование лексической компетенции студентов на основе работы с языковыми корпусами материала // Ученые записки Санкт-Петербургского университета технологий управления и экономики. 2022. Т. 23. № 2. С. 108–115. <https://doi.org/10.35854/2541-8106-2022-2-108-115>

Original article

Enhancing lexical competence on the basis of corpus linguistics at institutions of higher learning

Yuliya G. Davydova

St. Petersburg University of Management Technologies and Economics,
St. Petersburg, Russia, alexyulin@mail.ru

Abstract. The article deals with enhancing lexical competence of the students on the basis of modern language corpora. The author defines the notion ‘corpus’ in research, summarizes the views on importance of using corpora for linguistic and pedagogical purposes, analyzes the data of three modern English language corpora for improving lexical knowledge of the

© Давыдова Ю. Г., 2022

students. The article contains a complex of exercises which enables students to enhance lexical competence in semantics, collocations and pragmatics.

Keywords: corpus, pedagogical potential of using corpora, problem teaching, lexical competence, a complex of exercises

For citation: Davydova Yu. G. Enhancing lexical competence on the basis of corpus linguistics at institutions of higher learning. *Uchenye zapiski Sankt-Peterburgskogo universiteta tekhnologii upravleniya i ekonomiki. 2022;23(2):108-115. (In Russ.).* <https://doi.org/10.35854/2541-8106-2022-2-108-115>

Различные цифровые технологии (корпуса языкового материала, цифровые компоненты учебно-методического комплекса (УМК), цифровые образовательные ресурсы) в значительной степени способствуют формированию и совершенствованию всех компонентов лексической компетенции студентов, поскольку предоставляют обучаемому возможность получить доступ к новейшей лингвистической информации и проанализировать ее в соответствии с поставленными целями.

В настоящей статье рассмотрим возможности работы с корпусами языкового материала с целью совершенствования лексической компетенции студентов вузов. Д. Кристал в фундаментальной работе «Кембриджская энциклопедия английского языка» полагает, что «появление компьютерных баз данных языкового материала стало настоящей технологической революцией, кардинально изменившей традиционные подходы к сбору языкового материала и составлению словарей со времен Джонсона» [1, с. 483]. Под лингвистическим или языковым корпусом текстов в науке понимают «большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [2, с. 5].

По определению А. А. Шунейко, корпусная лингвистика — комплексная научная дисциплина, которая использует компьютерные технологии для выяснения природы языка, оптимизации коммуникации и решения иных лингвистических задач, а параллельно создает объекты, помогающие решать эти задачи, — корпуса данных. В процессе этой двунаправленной деятельности корпусная лингвистика углубляет и трансформирует наши представления о языке и коммуникации, помогает нам эффективнее взаимодействовать с техникой и друг с другом [3, с. 11].

Западные исследователи под корпусом понимают электронную коллекцию примеров устного и письменного английского языка, позволяющую ученым ответить на вопросы о лексическом составе, грамматике и прагматике языковых единиц [4]. Как российские, так и зарубежные ученые отмечают несомненную значимость корпусов для развития лингвистических исследований и их широкого применения в прикладной лингвистике и лингводидактике. Так, Д. Кристал пишет о том, что качественный корпус позволяет исследователям решать следующие задачи [1, с. 438]:

- 1) создавать более объективные модели функционирования языка, чем те, которые возможны на основе интроспекции;
- 2) делать выводы о частотности употребления языковых форм в языке в целом и в его вариантах;
- 3) анализировать все лингвистические черты любого текста в корпусе;
- 4) выдвигать гипотезы о функционировании языка;
- 5) объединять усилия различных групп исследователей в процессе анализа лингвистических проблем, выявленных в отдельном корпусе языкового материала.

Ученые обращают внимание на особую значимость корпусов в лингвистике и лингводидактике. Так, В. А. Плунгян утверждает, что «корпус не просто нужен, а необходим — профессиональным лингвистам, тем, кто имеет дело с фактами языка, а значит, должен эти факты собирать и систематизировать». При этом автор подчеркивает, что корпус — нечто большее, чем просто техническое подспорье. Это — фактически справочно-информационная система по русскому языку (любому языку, представленному в корпусе. — Ю. Д.), позволяющая получать ответы на самые неожиданные вопросы. Более того, позволяющая ставить новые проблемы, которых лингвистика прошлого почти не касалась» [5, с. 84].

Не меньшую значимость корпус имеет в лингводидактике. По мнению В. А. Плунгяна, для людей, не владеющих русским языком в качестве иностранного, корпус оказывается поистине незаменимым инструментом. Иностранец в высшей степени нуждается в инструменте, открывающем ему максимально широкий (и максимально комфортный) доступ в мир русского языка ввиду того, что его языковое сознание иное. Ничего лучше корпуса современная наука в этом случае предложить не может [5, с. 86].

Важность использования корпусов в лингводидактике определена и следующими факторами [3, с. 22]:

- 1) большой объем разнообразных данных создает гарантию того, что в совокупности корпуса фиксируют не случайные, а типичные явления, дают возможность воспринимать язык в широком диапазоне его компонентов и функций;
- 2) информация в корпусе содержится в своем естественном виде, она не препарирована по тем или иным причинам, не вырвана из контекста. Это создает предпосылки для ее объективного изучения. На степень объективности может накладывать ограничения разметка, но они вполне преодолимы;
- 3) корпус предполагает многократное и многоцелевое использование. Его создание — только начало исследовательской практики. Пополняемые или непополняемые массивы данных служат поколениям ученых с учетом того, что первому корпусу уже более полувека.

Значимым представляется утверждение А. А. Шунейко о том, что лингвистические корпуса в лингводидактике можно использовать в рамках двух подходов к обучению языку: репродуктивного и проблемного [3, с. 91]. Наиболее продуктивны, по мнению ученого, корпуса в рамках проблемного (когнитивно-интеллектуального) подхода к обучению языку, который ориентируется на активизацию познавательных функций человека. Реализуя проблемный подход, ученики, используя огромные массивы текстов, становятся первооткрывателями, создателями и соавторами языковых правил и закономерностей. Приобретенные на базе проблемного обучения знания эффективно и надолго усваиваются [3, с. 91].

К самым известным корпусам английского языка относятся Британский национальный корпус — *the British National Corpus* (100 млн слов); Международная база лингвистических данных Бирмингемского университета — *Collins Birmingham University International Language Database, COBUILD* (4,5 млрд слов); Корпус современного американского английского — *Corpus of Contemporary American English, COCA* (1,1 млрд слов); Корпус на основе глобальной сети — *the English Web Corpus* (38 млрд слов); Кембриджский корпус английского языка — *Cambridge English Corpus*; Кембриджский ученический корпус — *Cambridge Learner Corpus* (3 млн слов из международных экзаменов).

Попытаемся раскрыть возможности совершенствования лексической компетенции студентов на основе работы с тремя корпусами языкового материала

(the English Web Corpus [6], the British National Corpus [7] и Cambridge Learner Corpus [8]), отражающими современное состояние английского языка. Последняя версия корпуса the English Web Corpus включает в себя 38 млрд слов, представляющих лингвистически ценный материал. Пользователь может получить всестороннюю информацию о сочетаемости искомой лексической единицы (*word sketch*); синонимах и словах, входящих в лексико-семантическую группу (*thesaurus*); ключевых словах (*keywords*); списках лексических единиц, организованных по принципу частотности (*word lists*, например, все прилагательные, начинающиеся на а — *afraid, aghast etc.*); лексических единицах, включающих в себя несколько элементов (*n-grams*), исчерпывающую информацию о примерах употребления лексической единицы в контексте (*concordance*). В рамках данного корпуса пользователь получает доступ к ценному для лингводидактических целей учебному корпусу Cambridge Learner Corpus, содержащему материал международных экзаменов, проводимых Кембриджским университетом, таких как FCE, CAE и CPE. Материалом служит 2,9 млн слов, включенных в корпус из работ десяти тысяч студентов из более 60 стран, для которых английский язык является иностранным.

Британский национальный корпус (BNC) содержит 100 млн слов, собранных из письменных и устных источников, представляющих состояние английского языка, начиная с конца XX столетия. Письменная часть корпуса (90 %) составлена на базе широкого спектра источников (региональных и национальных газет, журналов, опубликованных и неопубликованных писем, академической и художественной литературы, школьных и университетских письменных работ). Устная часть (10 %) состоит из транскриптов неформальных разговоров (записанных добровольцами различных возрастов, регионов, социальных групп) из различных контекстов (деловых встреч, бесед в Парламенте, радишоу, телефонных звонков и др.). К характерным чертам Британского национального корпуса относятся [9]:

- 1) монолингвальность (корпус отражает состояние современного британского варианта английского языка);
- 2) синхрония (корпус описывает язык, начиная с конца XX столетия, а не его историческое развитие);
- 3) общий характер (представленность различных стилей и вариантов, тематических областей, регистров, не ограничен определенной областью).

Проведенный анализ языкового материала, входящего в корпуса, позволяет заключить, что корпуса способствуют значительному расширению следующих видов лексических знаний, составляющих лексическую компетенцию студентов:

- 1) знание о частотности употребления лексической единицы;
- 2) знание о типичных коллокациях;
- 3) знание об употреблении слова с учетом регистра и контекста;
- 4) знания о синонимах из определенной лексико-семантической группы;
- 5) знания о контекстуальном окружении лексической единицы.

Например, студент может составить карту семантического поля с центральным элементом *famous*. Согласно печатному словарю *Oxford Thesaurus* и его электронной версии, в синонимический ряд лексических единиц с центральным компонентом *famous* входят такие единицы, как *famous, historic, well-known, prominent, public, legendary, renowned, famed, glorious, celebrated* [10, p. 277]. Корпус the English Web Corpus предоставляет возможность ознакомиться с 1 000 лексических единиц, входящих в данное поле.

BNC и the English Web Corpus также содержат информацию о частотности употребления в корпусе всех лексических единиц, входящих в определенное семантическое поле. Так, согласно корпусу BNC, *famous* насчитывает 6 293 случая употребления в корпусе, *historic* — 2 266, *well-known* — 1 496, *prominent* — 2 231,

public — 38 032, *legendary* — 645, *renowned* — 659, *famed* — 267, *glorious* — 1 062, *celebrated* — 1 674 [7]. Таким образом, у студента существует возможность установить наиболее частотные и ценные для использования в определенном контексте лексические единицы и всесторонне их проанализировать.

Корпус *BNC* также позволяет определить частотность той или иной единицы в различных жанрах. Среди них — *spoken, fiction, magazine, newspaper, non-academic, academic* разговорный, художественная литература, журналы, газеты, неакадемический, академический язык. Например, в отношении лексемы *famous* приходим к выводу о том, что наибольшей частотностью данное слово обладает в журнальном и газетном регистрах языка [7], как видно из таблицы 1.

Таблица 1

Частотность лексемы *famous* в различных регистрах

SECTION	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	NON-ACAD	ACADEMIC	MISC
FREQ	6 293	255	710	838	924	898	539	2 129
WORDS (M)	100	10.0	15.9	7.3	10.5	16.5	15.3	20.8
PER MIL	62.93	25.59	44.63	115.40	88.28	54.44	35.16	102.18
SEE ALL SUB-SECTIONS AT ONCE								

Корпуса предоставляют широкие возможности углубить знания о **лексической сочетаемости** представленных в корпусе лексем (несколько компонентов до и после соответствующего слова). Так, в отношении *famous* (согласно *BNC*) уточняем, что наиболее типичными **существительными**, встречающимися после соответствующей лексической единицы, являются следующие [7]:

- *famous names* 85; *famous people* 61, *famous victory* 30; *famous example* 29; *famous Book* 28;
- *historic buildings* 173; *historic Scotland* 103; *historic cost* 51; *historic interest* 51; *historic houses* 50.

Еще более широкие возможности для анализа сочетаемостных моделей соответствующей лексической единицы предоставляет корпус *the English Web Corpus*. Так, в разделе *Word Sketch* отражены табличные данные следующих моделей сочетаний для прилагательного *famous* с соответствующими примерами [6]:

- 1) modifiers of famous;
- 2) nouns modified by famous;
- 3) famous and/or;
- 4) prepositional phrases;
- 5) infinitive objects of famous;

- 6) verbs complemented by famous;
- 7) verbs before famous;
- 8) subjects of 'be famous'.

Корпуса открывают и возможность сравнить частотность коллокаций для лексических единиц из одного синонимического ряда. Например, сравнить прилагательные *famous* и *legendary* предлагаем с помощью сопоставительной таблицы 2 для сочетаний (*collocations*) с существительными [7].

Таблица 2

Сравнение сочетаемости лексем *famous* и *legendary* с существительными

Word	Word 1	Word 2	Word 1 / Word 2	Score	Word	W2	W1	W2 / W1	Score
<i>people</i>	61	1	61.0	6.3	<i>status</i>	6	0	12.0	117.1
<i>victory</i>	30	0		6.1	<i>reputation</i>	4	0	8.0	78.1
<i>book</i>	28	0		5.7	<i>home</i>	3	0	6.0	58.5
<i>phrase</i>	25	0		5.1	<i>tales</i>	3	0	6.0	58.5
<i>names</i>	85	2		4.4	<i>blues</i>	3	1	3.0	29.3

Очень важной составляющей работы с корпусом в целях совершенствования лексической компетенции является всесторонний анализ контекста употребления лексических единиц (раздел *Concordance*). Как правило, студент анализирует значение соответствующей лексической единицы, обнаруживая ее в звучащем или письменном тексте, при обращении к толковому одноязычному словарю либо знакомстве со списком лексических единиц, сгруппированных вокруг соответствующей лексики в семантическом поле или тематической группе. Однако информация о всестороннем актуальном использовании лексики носителями языка в различных регистрах и контекстах остается за пределами анализа. Раздел *Concordance* в корпусах языкового материала позволяет решить данную задачу. Особенно ценна возможность изучить функционирование в реальном контексте не только отдельной единицы, но и многокомпонентных сочетаний.

С целью совершенствования лексической компетенции преподаватель может предложить студентам разнообразные задания, основанные на корпусах языкового материала. Приведем примеры таких заданий.

1. *Study the corpus and write the most common verbs found in the pattern verb + conclusion (s). Define the most frequent sphere of usage of the noun conclusion.* Изучите корпус и выпишите глаголы, встречающиеся в модели *verb + conclusion (s)*. Установите регистр, в котором существительное *conclusion* встречается наиболее часто.

2. *Study the corpus and write the most common nouns used with the verbs: make/do. Make up your own sentences with the corresponding collocations.* Изучите корпус и выпишите наиболее частотные существительные, используемые с глаголами *make/do*. Составьте предложения с соответствующими сочетаниями.

3. *Study the context of the words tactic and strategy. Define the difference between these words.* Изучите контекст употребления существительных *tactic* и *strategy*. Определите семантику существительных по контексту.

4. *Study the context (examples) and find the missing word.* Изучите таблицу 3 и вставьте слово, подходящее по контексту.

Таблица 3

Фрагмент конкорданса для существительного *decision* из BNC

my money was in sensible hands. Last week the society took the incredibly high-handed to refuse a cheque for 50,000 because it came from the profits of the famous
headed for Britain. And his parents back home in Fiji were right behind her’ They said they would not allow him to come otherwise so here I
jobs crisis, the Mellor crisis... and a crisis of leadership. Major’s to rely on a press resume written by others, and so avoid personally confronting
the first I’ve heard of it.’ Next day, we reported the and carried a Mirror Comment denouncing it. Now the CPS and Commander Penrose took

5. Study concordance lines of the noun *precaution* and write down the most typical collocations. Write your story with some of them. Изучите контекст употребления существительного *precaution*. Составьте рассказ, используя несколько словосочетаний.

6. Compare synonyms for the noun class from the *Thesaurus (Dictionary)* and semantic group from the *Corpus*. Define the frequency of the corresponding words from the *Thesaurus (Dictionary)*. Сопоставьте синонимы для существительного class из тезауруса (словаря) и корпуса. По корпусу определите частотность лексических единиц, входящих в словарь.

7. Study *Open Cambridge Learner Corpus* and find all the adjectives beginning with *a-* used by the students in the exams. Find out which adjectives are the most frequent ones. Study concordance lines and specify the context of their use and the exam. В корпусе *Open Cambridge Learner Corpus* найдите все прилагательные, начинающиеся на *a*, использованные студентами в кембриджских экзаменах. Установите наиболее частотные прилагательные и определите контекст их употребления.

Таким образом, анализ современных корпусов языкового материала позволяет сделать вывод о том, что работа с лексическими единицами корпуса может способствовать значительному углублению всех видов лексических знаний студентов (о семантике, частотности, сочетаемости, семантических и тематических связях, прагматике) лексических единиц, входящих в лексическую компетенцию студентов вузов.

Список источников

1. Crystal D. The Cambridge Encyclopedia of the English Language. Cambridge, 2002. 490 p.
2. Захаров В. П., Богданова С. Ю. Корпусная лингвистика: учебник. СПб.: Санкт-Петербургский государственный университет, Филологический факультет, 2013. 148 с.
3. Шунейко А. А. Корпусная лингвистика: учебник для вузов. М.: Юрайт, 2021. 222 с.
4. Better Learning Our Insights // Cambridge University Press. URL: <https://www.cambridge.org/gb/cambridgeenglish/better-learning-insights/corpus> (дата обращения: 01.12.2021).
5. Грудева Е. В. Корпусная лингвистика: учеб. пособие. М.: Флинта, 2012. 165 с.
6. EnTenTen: Corpus of the English Web // Sketch Engine. URL: <https://www.sketchengine.eu/ententen-english-corpus/> (дата обращения: 02.12.2021).
7. British National Corpus (BNC). URL: <https://www.english-corpora.org/bnc/> (дата обращения: 02.12.2021).
8. Open Cambridge Learner Corpus (Uncoded) // Sketch Engine. URL: <https://www.sketchengine.eu/cambridge-learner-corpus/> (дата обращения: 02.12.2021).
9. British National Corpus. URL: <http://www.natcorp.ox.ac.uk/> (дата обращения: 02.12.2021).
10. Oxford Learner’s Thesaurus. A Dictionary of Synonyms. Oxford: Oxford University Press, 2008. 1008 p.

Информация об авторе

Ю. Г. Давыдова — кандидат педагогических наук, доцент, доцент кафедры лингвистики и переводоведения.

Information about the author

Yu. G. Davydova — PhD in Pedagogy, Associate Professor, Associate Professor of the Department of Linguistics and Translation Studies.

Конфликт интересов: автор декларирует отсутствие конфликта интересов, связанных с публикацией данной статьи.

Conflict of interest: the author declares no conflict of interest related to the publication of this article.

Статья поступила в редакцию 14.12.2021; одобрена после рецензирования 16.03.2022; принята к публикации 08.06.2022.

The article was submitted 14.12.2021; approved after reviewing 16.03.2022; accepted for publication 08.06.2022.