

Карамнов Антон Сергеевич

**КОЛИЧЕСТВЕННАЯ ОЦЕНКА ПОВТОРЯЕМОСТИ И СЛОЖНОСТИ ЛЕКСИКИ В КОРПУСЕ
УЧЕБНИКА АНГЛИЙСКОГО ЯЗЫКА**

В статье освещается процедура проведения лингвистической оценки учебника английского языка на основе корпусного подхода. В центре внимания оказывается лексическая наполненность учебной книги. Анализ языкового материала осуществляется по признакам повторяемости активного словаря и соответствия уровню владения языком. Выводы исследования подкрепляются соответствующими количественными показателями.

Адрес статьи: www.gramota.net/materials/2/2014/6-1/21.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2014. № 6 (36): в 2-х ч. Ч. I. С. 82-86. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2014/6-1/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: voprosy_phil@gramota.net

LEXICOGRAPHICAL MODELING OF MODERN ENGLISH ECONOMIC TERMINOLOGY IN ELECTRONIC DICTIONARIES

Kameneva Nataliya Aleksandrovna, Ph. D. in Economics
Moscow State University of Economics, Statistics and Informatics
n-kameneva@yandex.ru

The article is devoted to investigating the basics of lexicographical modeling of the English and Russian economic terminology. The paper analyzes the difficulties under implementation of lexicographical modeling connected with the phenomena of polysemy, homonymy, lexical borrowings, necessity for elimination of lacunar phenomena and the presence of culture-specific vocabulary. The author accentuates that modeling as a process of harmonization of vocabulary and generation of multilingual texts promotes the establishment and development of cross-cultural communication, interlingual professional communication, international cultural and economic relationships.

Key words and phrases: lexicographical product; electronic dictionary; descriptive approach; terminological unit; term area; concept-based notion; lacunar phenomena; loan translation; borrowing; culture-specific vocabulary.

УДК 372.881.111.1

Педагогические науки

В статье освещается процедура проведения лингвистической оценки учебника английского языка на основе корпусного подхода. В центре внимания оказывается лексическая наполненность учебной книги. Анализ языкового материала осуществляется по признакам повторяемости активного словаря и соответствия уровню владения языком. Выводы исследования подкрепляются соответствующими количественными показателями.

Ключевые слова и фразы: корпусный подход; языковой корпус; учебник английского языка; лексическая наполненность; повторяемость лексики; количественный анализ.

Карамнов Антон Сергеевич

Тамбовский государственный университет имени Г. Р. Державина
k.a.s.89@mail.ru

КОЛИЧЕСТВЕННАЯ ОЦЕНКА ПОВТОРЯЕМОСТИ И СЛОЖНОСТИ ЛЕКСИКИ В КОРПУСЕ УЧЕБНИКА АНГЛИЙСКОГО ЯЗЫКА[©]

Постановка проблемы

Использование учебника в обучении иностранному языку рассматривается многими исследователями как необходимое условие достижения предметных и метапредметных результатов [6; 10; 11]. Этим обстоятельством обусловлена возрастающая потребность в обеспечении качества экспертной оценки учебников и учебных пособий по иностранному языку.

Анализ научной литературы показывает, что методический, дидактический и психологический аспекты экспертной деятельности в отношении учебных изданий раскрыты достаточно полно [1-3]. Созданный в рамках этих исследований инструментарий качественной оценки позволяет с высокой степенью вероятности определять, насколько учебник отвечает предъявляемым к нему требованиям в контексте обозначенных аспектов.

Вместе с тем недостаточно разработанной представляется методика оценки собственно языковой стороны учебных пособий по английскому языку, так как основным критерием отбора лексического материала по-прежнему остается интуиция и профессиональные воззрения авторского коллектива.

Корпусные технологии экспертной оценки языка учебников и учебных пособий по английскому языку позволяют провести количественный анализ их лексической наполненности, способствовать, таким образом, получению более объективной языковой картины и служить основанием для внесения изменений в издание с целью устранения выявленных недочетов.

Определение понятий

Корпусная лингвистика, изучающая язык посредством анализа его фактических аутентичных употреблений, в узком смысле является наукой о том, как создавать языковые корпуса и работать с ними.

Исследования, проводимые на материале больших массивов данных, подразумевают два вида интерпретации полученных результатов – количественный и качественный. Количественное изучение и качественное объяснение явлений в языковом корпусе составляет сущность количественных лингвистических исследований [5].

Языковой корпус представляет собой объемное, хранящееся в электронном виде собрание естественных случаев употребления языка в виде текстов различных жанров и стилей. Языковое наполнение корпуса никогда не случайно и варьируется от его объема и целей его конструирования.

Отличительной особенностью современных языковых корпусов является их машинно-считываемая форма. Это означает, что содержание корпуса может быть проанализировано с помощью автоматических средств компьютерной обработки данных. Программа, которая представляет собой систему управления языковым корпусом, то есть позволяет пользователю быстро находить и извлекать нужную информацию из корпуса, называется **корпусным менеджером**. Кроме этого, большинство корпусов, доступных онлайн, снабжены своими встроенными системами навигации и извлечения информации из массива текстовых данных.

Применение количественных методов для анализа языковых явлений предполагает несколько иной ракурс морфологической интерпретации слов как составляющих языкового корпуса.

Так, группа слов, образованная в результате флективных изменений (например, *worry – worries – worried – worrying*) называется **леммой** (lemma), в то время как группа слов с общей корневой основой (например, *manage – manager – management*) считается **семьей**, или гнездом, **слов** (word family).

В целях оптимизации автоматической обработки данных тексты корпуса, содержащие леммы и семьи слов, подвергаются токенизации и лемматизации.

Токенизация представляет собой автоматический процесс конвертирования всего текстового массива в отдельные токены (tokens) путем разделения кратких форм (*we've → we + 've*), отделения знаков препинания и удаления прописных букв.

Лемматизация как форма автоматического аннотирования заключается в сведении слов корпуса к соответствующим им лексемам, например, *he → he, could → can, brought → bring* и т.д.

В итоге корпус текстов может рассматриваться как совокупность **токенов**, т.е. отдельно стоящих лексических единиц, или как совокупность **типов** (types), т.е. неповторяющихся лексических единиц. Другими словами, если в тексте корпуса встречается три лексические единицы *'pretending'* в разных предложениях, они расцениваются как три токена или один тип.

Содержание и методология исследования

Предметом настоящего исследования является извлечение и изучения количественных языковых данных на уровне лексики и материала корпуса учебника английского языка.

Исходя из необходимости достижения цели, заключающейся в лингвистической оценке лексической наполненности учебника английского языка, задачи данного исследования включают:

- определение степени повторяемости активного словаря в корпусе учебника;
- проверку лексической наполненности корпуса учебника на предмет соответствия уровням владения иностранным языком.

Источником фактического материала для создания экспериментального корпуса послужил учебник английского языка для школьников среднего звена общеобразовательных учреждений с начальным уровнем владения языком. Выбранный курс был опубликован в 2007 году. Содержание учебника включает в себя десять модулей и соответствует уровню A1 по общеевропейской шкале.

Подготовительный этап исследования включал в себя процедуру создания корпуса учебника иностранного языка и составление референтных списков слов, с которыми в дальнейшем предполагалось проводить сравнительный анализ. Отметим, что основу корпуса составили собственно тексты и упражнения из Книги для студента. Формулировки учебных заданий, аппарат организации усвоения и аппарат ориентировки не учитывались при создании корпуса (подробнее см.: [4]).

Процессуальный этап состоял из ряда шагов по установлению количественных соответствий между лексическим содержанием обучения, заложенным в учебнике, и сгенерированными референтными списками слов. Кроме того, сравнительный анализ проводился и между частями корпуса учебника, соответствующими лексическому содержанию его модулей.

Источником референтного материала послужил список слов BNL (Billuroğlu-Neufeld-List), который на сегодняшний день является наиболее репрезентативным и сбалансированным перечнем общеупотребительной и академической лексики [8], а также лексическое содержание обучения в соответствии с общеевропейскими уровнями владения английским языком [9].

В качестве корпусного менеджера, инструмента количественной обработки текстовых данных и проведения сравнительного анализа была использована онлайн программа *TextLexCompare* [7].

Заключительный этап исследования был посвящен систематизации и интерпретации полученных данных, а также подведению итогов и выработке рекомендаций по совершенствованию лексической наполненности учебника английского языка.

Результаты исследования

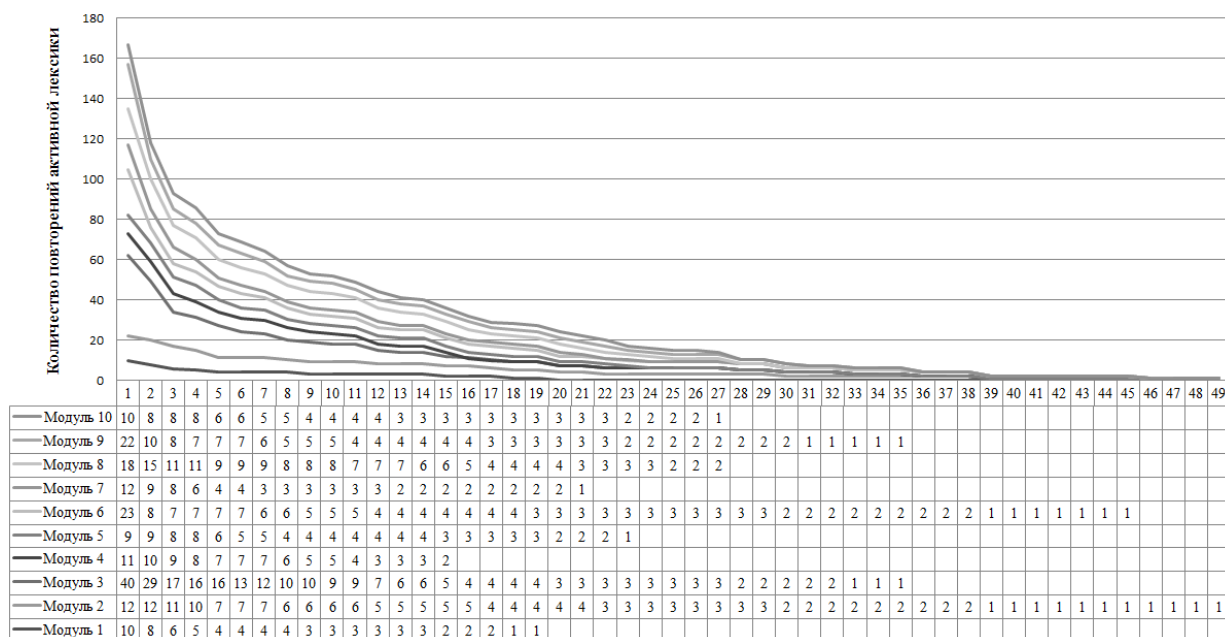
Рассмотрим, какие результаты были получены в ходе проведения лингвистической экспертизы корпуса учебника английского языка по двум направлениям анализа: степени повторяемости активного словаря и соответствию лексической наполненности уровням владения иностранным языком.

Для осуществления первого шага в процедуре анализа лексической наполненности в каждом модуле учебника был выделен активный словарь, состоящий из лексических единиц, предназначенных для продуктивного использования.

Определение степени повторяемости активной лексики в корпусе учебника проводится в две стадии. На первой стадии подсчитывается количество токенов для каждого модуля отдельно. Чтобы увидеть, сколько раз слово повторяется в рамках соответствующего тематического модуля в учебнике, обратимся к Диаграмме 1.

Диаграмма 1.

Степень повторяемости активной лексики в корпусе учебника



Количественные данные, отображенные на диаграмме, показывают, что, например, в первом модуле учебника было выделено 19 словарных единиц для активного усвоения. Три из них повторяются в модуле десять, восемь и шесть раз соответственно, в то время как есть две лексические единицы, которые упоминаются только один раз при первичном ознакомлении с ними.

Кроме этого, важно обратить внимание на то, что само число лексических единиц, отобранных для продуктивного использования, часто сильно различается от модуля к модулю. Так, если в четвертом и седьмом модулях учебника таких единиц 15 и 21 соответственно, то в шестом и втором модулях их количество возрастает до 45 и 49 в указанном порядке.

На второй стадии определения степени повторяемости активной лексики в корпусе учебника устанавливается количество повторений словарных единиц предыдущих модулей в корпусе последующих модулей. Обратимся к Таблице 1.

Таблица 1.

Степень повторяемости активной лексики по модулям в корпусе учебника

	Количество типов по модулям	Количество повторяющихся типов	Примеры типов с наибольшим числом повторений
Модуль 1	19	–	
Модуль 2	49 (68)	3	English (2)
Модуль 3	35 (103)	8	big (8) small (7) beautiful (5)
Модуль 4	15 (118)	23	hair (6) fair (5) French (5) guitar (3)
Модуль 5	23 (141)	28	bear (9) leg (9) long (7) tall (6)
Модуль 6	45 (186)	28	mum (5) scarf (5) dad (4) brother (3)
Модуль 7	21 (207)	32	Spain (9) head (6) big (5) watch (4)
Модуль 8	27 (234)	38	American (6) kitchen (4) March (4) small (4)
Модуль 9	35 (269)	44	book (5) boot (5) painting (5) tomato (5)
Модуль 10	27 (296)	45	summer (8) wear (7) hot (6) travel (6)

В столбце «Количество типов по модулям» указано число единиц активной лексики, отобранной в каждом модуле учебника, с дополнительным обозначением в скобках суммарной количества этих единиц с учетом предыдущих показателей. Это означает, что если в третьем модуле учебника 35 выбранных лексических единиц, то в целом в корпусе первых трех модулей учебника 103 таких единицы.

Третий и четвертый столбцы таблицы отражают количество повторяющихся типов активной лексики в модулях учебника с их наиболее частотными примерами. Так, например, в третьем модуле встречается 8 из 68 единиц активной лексики, составляющей содержание обучения в первых двух модулях.

Обратимся теперь ко второму направлению оценки – проверке лексической наполненности корпуса учебника на предмет соответствия уровням владения иностранным языком. Для этого его содержание сравнивается со списками слов по уровням А1-С2 в соответствии с общеевропейской шкалой. Полученные данные представлены в Таблице 2.

Таблица 2.

Лексическая наполненность корпуса учебника по уровням A1-C2

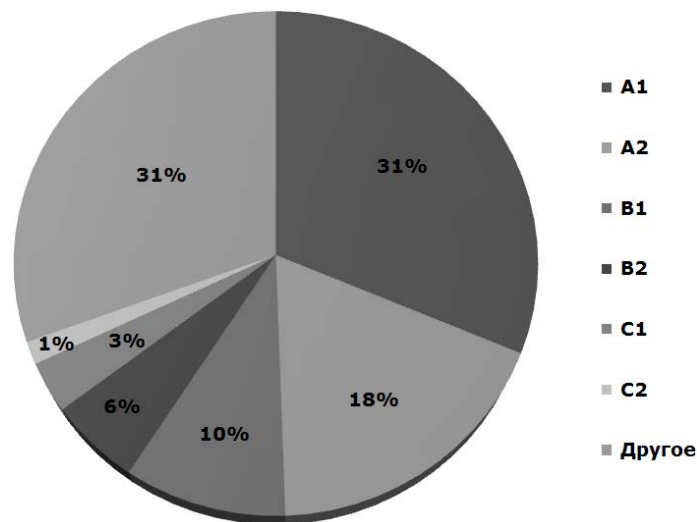
Уровень	A1	A2	B1	B2	C1	C2
Число типов по уровням	1137	1368	1622	1839	2128	1799
Число соответствий в корпусе	755	438	235	134	79	35

Из Таблицы видно, что в корпусе учебника английского языка содержится лексика всех уровней сложности от A1 до C2. Основная часть корпуса представлена общеупотребительной лексикой уровней A1 (755 типов) и A2 (438 типов). Отметим, что общее количество типов в самом корпусе составляет 2413 единиц.

Чтобы полнее представить картину лексической наполненности корпуса анализируемого издания, изобразим ее в виде круговой диаграммы (см. Диаг. 2).

Диаграмма 2.

Лексическая наполненность корпуса учебника



Процентные данные диаграммы показывают, что треть корпуса учебника иностранного языка составляет лексика, которая не входит в словарный состав ни одного из уровней лексического содержания обучения в соответствии с общеевропейской шкалой владения иностранным языком.

На долю этого словарного блока приходятся имена собственные, некоторые терминологические единицы, а также пласт менее частотной лексики.

Обсуждение результатов и рекомендации

Количественный анализ корпуса учебника английского языка на предмет повторяемости и сложности лексики позволяет сделать следующие выводы.

В процессе определения степени повторяемости активного словаря в корпусе учебника было выявлено, что, во-первых, количество единиц лексики, предназначенной для продуктивного использования, сильно варьируется от модуля к модулю (от 15 до 49), и, во-вторых, только около 20% изученной лексики повторяется в последующем содержании корпуса учебника.

В связи с этим авторам учебной книги и тем, кто использует ее в образовательном процессе, может быть рекомендовано:

а) унифицировать количество единиц активной лексики в модулях учебника, сократив их список за счет типов с минимальной повторяемостью;

б) повысить показатель повторяемости активной лексики после ее первичного изучения в последующих модулях учебника путем точечного добавления требуемых словарных единиц по тематическим областям.

В направлении проверки лексической наполненности корпуса учебника на предмет соответствия уровням владения иностранным языком было выявлено, что издание на 38% состоит из лексики, уровень которой выше заявленного. Словарный состав, соответствующий уровню A1, соотносится лишь с 31% корпуса учебника. Еще 31% составляет лексика, находящаяся за пределами содержания обучения, обозначенного для уровней A1-C2.

Такие данные дают основание рекомендовать меры по упрощению лексического содержания корпуса учебника, которые могут заключаться в удалении лексики уровней C1-C2 и сведении к минимуму лексики уровней B1-B2. Кроме этого, может оказаться полезным сокращение малоупотребительной лексики в пользу повторения более частотных эквивалентов.

Список литературы

1. Арутюнов А. Р. Теория и практика создания учебника русского языка для иностранцев. М.: Русский язык, 1990. 168 с.
2. Бим И. Л. Методика обучения иностранным языкам как наука и теория школьного учебника. М.: Русский язык, 1977. 288 с.
3. Гельфман Э. Г., Холодная М. А. Психодидактика школьного учебника. Интеллектуальное воспитание учащихся. СПб.: Питер, 2006. 384 с.
4. Карамнов А. С. Модель создания корпуса учебника английского языка // Научный диалог. 2013. № 2 (14). С. 59-69.
5. Мильруд Р. П., Сысоев П. В., Кащеева А. В., Карамнов А. С. Введение в корпусную лингвистику: учеб. пособие. Тамбов: Издательский дом ТГУ им. Г. Р. Державина, 2012. 129 с.
6. Ansary N., Babaii E. Universal Characteristics of EFL/ESL Textbooks: A Step Towards Systematic Textbook Evaluation [Электронный ресурс] // The Internet TESL Journal. 2002. Vol. 8. № 2. URL: <http://iteslj.org/Articles/Ansary-Textbooks/> (дата обращения: 25.03.2014).
7. Cobb T. Computing the Vocabulary Demands of the L2 Reading // Language Learning & Technology. 2007. Vol. 11. № 3. P. 38-63.
8. Hancioglu N., Neufeld S., Eldridge J. Through the Looking Glass and into the Land of Lexico-Grammar // English for Specific Purposes. 2008. Vol. 27. P. 459-479.
9. Neufeld S., Eldridge J. LexiCLIL: a Lexical Syllabus for the Common European Framework for English [Электронный ресурс]. URL: <http://ru.scribd.com/doc/20386024/Common-English-Lexical-Framework> (дата обращения: 25.03.2014).
10. Noordin N., Samad A. A. Examining the importance of EST and ESL textbooks and materials: objectives, content and form [Электронный ресурс]. URL: http://www.esp-world.info/Articles_9/textbooks.htm (дата обращения: 25.03.2014).
11. Sheldon L. E. Evaluating ELT textbooks and materials // ELT Journal. 1988. Vol. 42/4. P. 237-246.

**QUANTITATIVE EVALUATION OF REPETITION AND COMPLEXITY
OF VOCABULARY IN THE CORPUS OF ENGLISH LANGUAGE TEXTBOOK**

Karamnov Anton Sergeevich

Tambov State University named after G. R. Derzhavin

k.a.s.89@mail.ru

The article considers the procedure of linguistic evaluation of an English language textbook on the basis of corpus approach. The special attention is paid to the lexical completion of a textbook. Analysis of language material is carried out according to the criteria of repetition of the active vocabulary and adequacy for the level of language competence. The conclusions of the research are supported by the appropriate quantitative indices.

Key words and phrases: corpus approach; linguistic corpus; English language textbook; lexical completion; repetition of vocabulary; quantitative analysis.

УДК 81'42; 801.7

Филологические науки

Статья посвящена определению роли паралингвистических средств в процессе коммуникации. Выявляются особенности использования таких средств в ситуации социального и ситуативного равенства и неравенства партнеров по коммуникации. Рассматриваются семантические особенности описаний паралингвистических средств в художественном тексте. Делается вывод о том, что паралингвистические средства являются более значимыми индексами социального статуса, чем вербальные средства.

Ключевые слова и фразы: паралингвистические средства; вербальные описания паралингвистических средств; социальный статус; социальное равенство (неравенство); ситуативное равенство (неравенство).

Кильмухаметова Елена Юрьевна, к. филол. н.

Национальный исследовательский Томский государственный университет

vadetecum72@mail.ru

**ПАРАЛИНГВИСТИЧЕСКАЯ ИНДИКАЦИЯ СОЦИАЛЬНОГО СТАТУСА
(НА ПРИМЕРЕ ФРАНЦУЗСКОЙ ХУДОЖЕСТВЕННОЙ ЛИТЕРАТУРЫ)[©]**

В последние десятилетия наблюдается усиление интереса к внеязыковым компонентам речевого общения, среди которых особое место отводится невербальным и паралингвистическим средствам общения. Паралингвистические средства включают в себя, по мнению одних, фонационные, кинесические и графические средства [2; 5], по мнению же других, исключительно фонационные [4], к которым относят параметры звучания речи: мелодика интонации, длительность пауз, темп, ритм, высота тона, громкость и др. [Там же]. В работе под паралингвистическими средствами понимаются голосовые характеристики говорящего субъекта.