

Министерство образования Республики Беларусь
Учреждение образования «Барановичский государственный университет»
Студенческое научное общество БарГУ

СОДРУЖЕСТВО НАУК. БАРАНОВИЧИ-2016

Материалы XII Международной
научно-практической конференции
молодых исследователей

(Барановичи, 19—20 мая 2016 года)

В трёх частях

Часть 1

Барановичи
БарГУ
2016

В часть 1 сборника материалов XII Международной научно-практической конференции молодых исследователей «Содружество наук. Барановичи-2016» включены материалы докладов по актуальным проблемам различных научных направлений: филологические науки (русский, белорусский языки); лингвистика (английский, немецкий языки); методика преподавания иностранного языка, правоведение.

Сборник адресован преподавателям и студентам учреждений высшего образования, магистрантам, аспирантам.

Редакционная коллегия:

А. В. Никишова (гл. ред.), А. В. Прадун, В. Н. Кременевская (отв. секретари),
О. Н. Фенчук, Т. М. Пучинская, А. А. Савко, Н. А. Егорова, М. В. Андрияшко

Рецензент

кандидат педагогических наук, доцент, заведующий кафедрой германских языков факультета международных отношений Белорусского государственного университета М. Ф. Арсентьева

ЧАСТОТНЫЕ СЛОВАРИ И ПРИНЦИПЫ ИХ СОСТАВЛЕНИЯ (НА ПРИМЕРЕ РУССКОЙ И АНГЛИЙСКОЙ ЛЕКСИКОГРАФИИ)

Введение. При изучении иностранного языка важно владеть информацией о повторяемости языковых единиц этого языка. Учёными было доказано, что в английском языке примерно в 90% письменного текста встречается от четырёх до пяти тысяч наиболее часто употребляемых слов. То же самое можно сказать и про устную речь. Очевидно, что данная информация будет полезна при работе с лексикой изучаемого языка. Рационально учить те слова, которые наиболее часто возможно услышать или прочитать в повседневных ситуациях. Узнать, как часто употребляется конкретное слово, можно в специальных лингвистических словарях, называемых частотными. На сегодня такие словари дают информацию не только об отдельных словах, но и о словосочетаниях, фразовых глаголах, идиомах.

Основная часть. Составлением, классификацией и изучением словарей занимается специальный раздел языкознания — лексикография. Существуют различные виды словарей: энциклопедические и лингвистические, общие и специальные, нормативные и дескриптивные, одноязычные и многоязычные.

Частотный словарь является новым видом специальных лингвистических словарей, который даёт информацию о частоте употребления слов. Первые словари такого рода появились во второй половине XX в. Однако не всегда частотные словари фиксируют лексику современного языка — может размещаться реестр слов определённой эпохи, лексемы определённого автора и т. п. Таким образом, необходимо быть внимательным при выборе частотного словаря. Так, например, в 1953 г. в Соединённых Штатах Америки Г. Г. Йоссельсоном был издан частотный словарь, но большая часть материала для его составления принадлежала лексемам американского варианта английского языка XIX в.

В русской лингвистической науке известен «Частотный словарь современного русского литературного языка» [1] Э. А. Штейнфельдта. Основное назначение издания — определение лексического минимума, включённого в школьные учебники русского языка. Процесс создания подобного словаря в середине XX в. был достаточно трудоёмким по причине ручной обработки информации. Сегодня подобные словари составляются при помощи вычислительной техники. На стыке лингвистики и информационных технологий возникла новая наука — компьютерная лингвистика, которая включает в себя одно из актуальных направлений в языкознании XXI в. — корпусную лингвистику. Эта область лингвистики занимается составлением корпусов текстов. Одним из направлений работы является использование этих корпусов для различных целей, в том числе и для лингвистических исследований. Под корпусом текста (обычно в электронной форме) в современной лингвистике понимается набор текстов, отобранных по определённым критериям, задачам, сферам использования [2, с. 167—169].

Для составления частотного словаря анализируются корпусы текстов. В зависимости от назначения словаря могут использоваться тексты определённого языка, тексты определённой области знаний, тексты конкретного автора или конкретных его произведений. Но обычно авторы при анализе стараются одновременно охватить максимально возможное количество текстов всех стилей. Средняя частота вычисляется по принципу употребления конкретных слов среди миллиона слов корпуса.

При составлении списков частоты употребления авторы сталкиваются с некоторыми проблемами. Во-первых, на другом аналогичном корпусе текста результат может быть различным. Во-вторых, на позицию слова в списке может повлиять частота употребления слова в одном конкретном тексте. Именно поэтому данные словари дают возможность сравнения нескольких корпусов. Следовательно, это позволяет определить, какие из слов характерны для каждого из этих корпусов в отдельности.

Цель данной работы — исследовать и сравнить частотные словари, созданные для английского и русского языков.

В рамках данного исследования производилось сравнение структуры и содержания двух словарей: «Частотного словаря современного американского английского языка» (“A Frequency Dictionary of Contemporary American English”) [3] и «Частотного словаря современного русского языка» [4].

При составлении двух словарей использовались тексты национальных корпусов. В этих корпусах в электронном виде собираются тексты всех жанров на национальном языке. Интересно, что анализу подвергались не только художественные и технические тексты, но и аудиозаписи «живых» разговоров, видеозаписи телепередач. Без внимания не остались и популярные журналы, газеты и веб-страницы. Так, при составлении словарей было проанализировано более 400 млн слов.

Говоря об общей структуре, оба этих словаря дают информацию о частоте повторения слов и предоставляют список наиболее употребляемых слов. В словарях лексемы представлены несколькими способами, соответственно, составлено несколько списков слов. В связи с этим данные словари, как правило, объёмны.

Первый способ представления информации в словаре — в алфавитном порядке. Достоинство алфавитной подачи в том, что есть возможность быстрого нахождения нужного слова, рядом с каждым словом приведена средняя частота его употребления. Недостаток заключается в отсутствии разделения по частям речи.

Второй способ — частотный. Здесь список слов построен в порядке ранжирования от наибольшей к меньшей частоте использования лексем. Данный способ удобен для быстрого нахождения статистических данных и апеллирования ими в нужной ситуации. Однако, как и в первом способе, лексемы не классифицируются по частям речи.

Третий способ подачи лексикографического реестра — по частям речи. Составлены списки наиболее распространённых имён существительных, имён прилагательных, глаголов, местоимений, наречий, числительных, служебных частей речи. Сложность составления подобного типа классификации наблюдается в английском языке, так как одни и те же формы слов могут выступать как в роли имён существительных, так и прилагательных. В таких случаях подобные лексемы в роли разных частей речи характеризуются разной частотностью употребления. Очевидно, что для лингвистических целей наиболее информативным является способ подачи материала в частотном словаре согласно частям речи.

В структуре словарей имеются некоторые отличия. В словаре русского языка шире представлена классификация слов. Имеются следующие разделы: словарь наиболее значимой лексики художественной литературы, словарь наиболее значимой лексики публицистики, газетно-новостной лексики, нехудожественной литературы, устной речи. Авторы попытались учесть, что в каком-либо функциональном стиле некоторые слова употребляются в большем количестве, чем в других стилях. Например, самым значимым (типичным) словом в публицистике является существительное «год», а в художественной литературе — глагол «сказать». Частотный список составлен также и для имён собственных, аббревиатур и букв алфавита, но построены они в алфавитном порядке. Этот словарь также позволяет узнать, в каком количестве текстов встречается определённое слово. Издание также приводит статистику, что слово русского языка состоит в среднем из пяти символов, а предложение состоит в среднем из десяти слов [5].

Имеются различия и в структуре словарной статьи. В русскоязычном словаре говорится частотность слова и к какой части речи оно относится (например: *быть* — глагол — 12844.9).

В «Частотном словаре современного американского английского языка» кроме частотности употребления приводятся и слова, с которыми наиболее часто сочетается рассматриваемое слово. Чтобы реализовать это, компьютерная программа проанализировала каждый контекст, в котором употреблено необходимое слово. Но результаты компьютера всё равно были проверены несколькими носителями английского языка (например: *be — verb — there, if, many, able, long, always, likely, often, available — 14338665 — 0,99*).

Из примера видно, что для ранжирования слов применяют характеристики, такие как частотность (“frequency”) и дисперсия (“dispersion”). Корпусы делятся на одинаковые по размеру части. Дисперсионное значение «1» указывает, что в каждой части слово имеет одинаковую частотность. Дисперсионное значение «0,10» будет означать, что в одних частях слово встречается часто, а в других может практически не встречаться [6].

Во время анализа словарей нами было проведено сравнение частоты употребления лексем русского и английского языков. Так, при детальном изучении наиболее используемых слов разных частей речи (в частотных списках под номерами от одного до десяти) оказалось, что в английском и русском языках они практически одни и те же, например, имена существительные: *человек — a person, год — a year, время — time, жизнь — life, день — a day*; глаголы: *быть — to be, сказать — to say, знать — to know, идти — to go, делать — to do, to make*; имена прилагательные: *новый — new, хороший — good, большой — big, последний — last, маленький — little, small*; местоимения: *я — I, он — he, это — it, этот — this*; предлоги: *в — in, на — on, с — with, для — for*; союзы: *и — and, но — but, как — as*.

Внимательная работа с частотными списками показывает, что для нормального общения достаточно примерно две тысячи слов в активном словарном запасе и ещё примерно столько же в пассивном. Наиболее часто употребляемые слова создают базу для дальнейшего более глубокого и специализированного изучения языка.

Заключение. Анализ лексикографических работ показал, что лингвисты, составляющие частотные словари различных языков, имеют одну и ту же цель — дать информацию о частоте употребления слова в языке. Различия в подаче материала в словарях обусловлены замыслом авторов, стилистической окраской лексики, жанровой принадлежностью текстов и спецификой использования лексем. Частотные словари можно использовать при преподавании как родного, так и иностранного языка. Эти словари применимы также при самостоятельном изучении языков, необходимы в научно-исследовательской деятельности вопросов современного языкознания.

Список цитируемых источников

1. Штейнфельдт Э. А. Частотный словарь современного русского литературного языка. Таллин, 1963.
2. Яскевич А. А. Корпусная лингвистика // Энцикл. для школьников и студентов : в 12 т. Минск : Белорус. энцикл., 2009. Т. 1 : Информацион. общество. XXI век / под общ. ред. В. И. Стражева. С. 167—169.
3. Davies M. A., Gardner D. Frequency Dictionary of Contemporary American English : word sketches, collocates, and thematic lists. London : Routledge, 2010.
4. Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М. : Азбуковник, 2009.
5. Там же. С. 1 065.
6. Davies M. A., Gardner D. Frequency Dictionary of Contemporary American English. P. 5.