

УДК 811.111'367.623'37(045)

*О.Ю. Буйнова***КОРПУСНАЯ ЛИНГВИСТИКА КАК ИНСТРУМЕНТ ИССЛЕДОВАНИЯ
ЯВЛЕНИЯ МНОГОЗНАЧНОСТИ**

В статье исследуется семантика прилагательного «big» с использованием Британского корпуса данных и выявляются особенности употребления этой полисемантической единицы; освещается роль комбинаторной лингвистики при анализе функционально-речевого аспекта языка и явления многозначности. Актуальность темы обусловлена тем, что высокие технологии помогают решить традиционные проблемы языкознания по-новому и их решение достигается легче и доказательнее. Соответственно, цель исследования – рассмотреть, каким образом корпусная лингвистика способствует выявлению многозначности слов и систематизации семантической структуры лексических единиц. В работе представлены основные способы обработки большого массива данных и контекстного метода решения проблемы полисемии слов; произведен семантический анализ конкретной лексической единицы с учетом типов значений и метода статистического подсчета; предлагается сопоставительный анализ материала в словарных статьях, представленных в электронных и печатных изданиях таких авторитетных лексикографических источниках, как Collins Cobuild Advanced Learner's Dictionary, Longman Language Activator, Dictionary of Linguistics and Phonetics, Oxford Advanced Learner's Dictionary. В итоге определяется степень семантизации вовлеченной в исследование лексической единицы «big» в рамках микро- и макроконтэкстов.

Ключевые слова: языковой корпус данных, комбинаторная лексикология, полисемия, лексико-семантическая сочетаемость, имена прилагательные широкой семантики.

DOI: 10.35634/2412-9534-2023-33-4-790-797

Эффективные методы применения информационных технологий в исследовании языковых процессов приобретают в настоящее время особую актуальность. Так, одним из наиболее перспективных направлений в современной теоретической и прикладной лингвистике, основанном на использовании языковых электронных корпусов и компьютерных технологий, является корпусная лингвистика. Согласно определению В.И. Подлесской «корпусная лингвистика – современная, быстро развивающаяся область, возникшая вследствие растущих потребностей науки о языке во внедрении компьютерных технологий для работы с большими массивами языковых данных» [5]. Корпусная лингвистика предлагает теоретические основы и практические механизмы создания и использования представительного объема данных, предназначенных для актуальных исследований в различных областях науки о языке.

Обратимся к ряду определений, предложенных современными отечественными и зарубежными лингвистами, для уточнения ключевого понятия «корпус текстов». Так, В.В. Рыков в работе «Корпусная лингвистика и информационный поиск» уточняет, что «корпус текстов – это определенным образом организованное множество, элементами которого являются тексты» [6, с. 18]. Более полное толкование этого понятия дает В.П. Захаров: «Под названием лингвистический корпус текстов понимается большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [4, с. 6]. По мнению одного из самых авторитетных современных британских лингвистов Дэвида Кристала, корпус – это набор лингвистических данных, представленных в письменной форме или в форме транскрипции устной речи, которые могут быть использованы как отправная точка лингвистических описаний или как средство подтверждения определенных гипотез о языке: a computer corpus is a large body of machine-readable texts. Increasingly large corpora (especially of English) have been compiled since the 1980s, and are used both in the development of natural language processing software and in such applications as lexicography, speech recognition, and machine translation [10, с. 117]. Безусловно, развитие современных интеллектуальных программных систем, предназначенных для обработки текстов, требует большой экспериментальной лингвистической базы.

Появление и развитие информационных технологий привело к тому, что появилась возможность изучать язык не только по традиционным «хранилищам» языковых данных, таким как словари, произведения художественной литературы, письменные и устные тексты различных жанров, но и вводить в компьютерную базу данных и обрабатывать большие массивы текстов. Этот качественный скачок особенно ценен для прикладной лингвистики, так как составление словарей является чрезвычайно важной

теоретической и практической работой, определяющей, по существу, состав и структуру всех лингвистических исследований. В этой связи нельзя не отметить и значимую ценность применения полученных результатов к практическим задачам, таким, например, как решение актуальных вопросов лингводидактики.

Кроме того, представительный массив языковых данных позволяет изучить динамику процессов изменения лексического состава языка за определённый период и провести сопоставительный анализ функционирования лексико-семантических единиц в разных языках. Именно современные методы исследования предлагают новые инструменты и дают возможность проверить гипотетические построения как методом обратной связи, так и точными количественными подсчётами. Соответственно, в настоящее время корпусная лингвистика, в рамках которой изучаются вопросы распределения лингвистических явлений в разных языках и объективным путем получают новые лингвистические данные, приобретает особое значение. Так, на основе полученных данных был создан национальный языковой корпус, значение которого для изучения языка в целом и его лексического состава сложно переоценить. Национальный корпус представляет собой собрание текстов в электронной форме, представляющих данный язык на определенном этапе его существования, во всём многообразии жанров, стилей, территориальных и социальных вариантов и т.п. Обязательное требование к национальному корпусу – его репрезентативность.

Британский национальный корпус (British National Corpus) – это большой представительный массив британского английского языка, по образцу которого создавались многие современные корпуса различных языков. Он включает метатекстовую и частеречную разметку, характеризуется использованием полных текстов и наличием подкорпуса устной речи объемом 10 млн., а объем всего корпуса насчитывает свыше 100 млн. словоупотреблений [8].

Целесообразность использования Британского национального корпуса определяется, прежде всего, его репрезентативным объемом, который гарантирует точность данных и обеспечивает полноту представления всего спектра языковых явлений, причем данные разного типа представлены в корпусе в определенном контексте, что способствует их объективному и детальному изучению. Подобный массив постоянно пополняется и расширяется, предоставляя исследователям возможность отслеживать как статичное состояние, так и динамичные языковые процессы. В области прикладного языкознания конкордансы получили особое признание лингвистов благодаря новым возможностям эффективного исследования языка и обработки материала текстов различного рода.

Эта специализированная система, располагающая программными средствами для поиска данных в корпусе, предлагает исчерпывающую статистическую информацию и представляет полученные результаты в удобном формате. Пользователь имеет доступ ко всем контекстным употреблениям, в которых встречается словоформа, причем существуют разнообразные возможности по извлечению из корпуса следующей информации: поиск конкретных словоформ, поиск словоформ по леммам, в виде разрывной или неразрывной синтагмы, по набору морфологических признаков, отображение информации об этимологии, типе текста и т.п.

Особую значимость представляет получение результатов поиска с указанием контекста заданной длины, получение статистических данных и сохранение отобранных строк конкорданса в отдельном файле. Результаты поиска оформляются в виде конкорданса, где искомая единица содержится в контекстном окружении в сопровождении конкретных статистических данных, при этом фиксируются частотные характеристики слова и дается его сочетаемость. Система позволяет настраивать формат выдачи информации: менять длину левого и правого контекста, задавать объем и порядок сортировки данных, лингвистические и экстралингвистические характеристики. Использование результатов Британского национального корпуса помогает успешно решить актуальные проблемы комбинаторной лексикологии, под которой понимается отдельный раздел комбинаторной лингвистики, имеющий собственный предмет описания – комбинаторно-синтагматические свойства слова.

В процессе изучения лексического состава языка синтагматические характеристики слов представляют особую важность. Нельзя не согласиться с автором монографии «Теоретические проблемы комбинаторной лингвистики. Лексикология. Лексикография» М.В. Влавацкой, которая отмечает, что «систематизация комбинаторных возможностей слова вносит значительный вклад в описание строя языка, является основанием для классификации слов по их синтаксическим и лексическим сочетательным свойствам и лежит в основе отражения когнитивных процессов в интерпретации языковых выражений» [2, с. 241].

Повышенное внимание к проблеме сочетаемости на современном этапе объясняется произошедшим сдвигом к исследованию функционально-речевого аспекта языка, который вскрыл целесообразность введения сочетаемости в «функционально-коммуникативное русло изучения синтагматики лексических единиц в целом» [1, с. 73].

Необходимо отметить, что целью комбинаторной лексикологии является изучение проблем, связанных с сочетаемостью лексических единиц, выявлением механизмов их взаимодействия, установлением закономерностей их связей, и, наоборот, отличий. Достижение данной цели сопряжено с изучением соотношения сочетаемости и семантики слова, описанием структуры синтагматического компонента значения слова, рассмотрением типов сочетаемости слов, выделяемых на основе различных принципов (по характеру сочетаемости, по способу связи между компонентами, по количеству сочетающихся слов и т. д.) [2, с. 242].

Для достижения поставленной цели настоящей работы необходимо обратиться к особому разделу комбинаторной лексикологии – комбинаторной семасиологии: в нём описываются семантические механизмы отношения значения слова и его сочетаемости. Более точно комбинаторную семасиологию можно определить как область лексикологии, содержанием которой является изучение семного состава лексического значения сочетающихся слов, выявление общих сем, позволяющих связывать слова в речевой цепи, а также рассмотрение структуры синтагматического компонента значения слова. Д.Н. Шмелёв утверждает, что значение слова и его сочетаемость находятся в отношении взаимозависимости. Слово, чтобы сочетаться с другой лексической единицей, уже должно обладать определённым смыслом, сочетаемость же заложена в самом слове и является важнейшим собственно языковым показателем данного значения [7, с. 242]. Эта идея находит свое подтверждение и в законе семантического согласования В.Г. Гака: в силу действия данной нормы сочетаемость слов оказывается мощным средством различения лексико-семантических вариантов полисемантической единицы [3].

Проблемы сочетаемости слова неразрывно связаны с его семантикой, на определение природы которой влияют различные факторы: структура лексического значения, языковые и экстралингвистические ограничения, идиоматичность и неидиоматичность лексических единиц, число соотносимых слов и их расположение. Комплексное рассмотрение данных проблем способствует детальному исследованию природы соотносительности лексических единиц, которые следует отнести к области комбинаторной семасиологии, основу которой представляет, с одной стороны, теория лексического значения слова, с другой – теория сочетаемости.

В целом, учитывая важность изучения комбинаторно-синтагматических свойств лексических единиц, которые играют основную роль в процессе производства речи, и сложность обозначенных проблем в плане теоретического осмысления, необходимо отметить как научную целесообразность, так и практическую значимость изучения синтагматики и комбинаторики слова.

Образование коллокаций обуславливается не только правилами морфо-синтаксической и лексико-фразеологической сочетаемости, но и языковым мышлением. Коллокация – это единица языка и единица речевой традиции. Синклер предположил «принцип идиомы» в использовании языка. «The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable to into segments. The mental lexicon of any native speaker contains single-word units as well as phrasal units or collocations. Mastery of both types is an essential part of the linguistic equipment of the speaker or writer and enables him to move swiftly and with little effort through his exposition from one prefabricated structure to the next» [13, с. 21].

Благодаря современным компьютерным технологиям, а именно на основе их активного использования в корпусной лингвистике, с помощью автоматизированного поиска и извлечения коллокаций и последующей их аналитико-синтетической обработки, позволяющей выявить весь спектр семантики того или иного слова на основе контекста, создаются самые авторитетные словари, в которых регистрируется максимально исчерпывающий перечень значений слов. Соответственно, на данном этапе развития науки о языке корпусная лингвистика располагает самыми оптимальными средствами решения проблемы полисемии слов.

Одним из основных методов разрешения многозначности является контекстологический метод семантизации слов как в широком, так и в узком контекстах. Слово изучается не в отрыве от основного контекста, а как его часть, что позволяет изучить слово как единицу коммуникации. Очевидно, что

одним из преимуществ электронного корпуса над печатными изданиями является презентация исследуемой лексической единицы в более широком контексте. Минимальным или, как его еще называют, узким контекстом лексической единицы является коллокация.

Материалом для исследования в работе послужило частотное прилагательное с широкой семантикой «big», обозначающее размер: синонимический ряд в русском языке представлен следующими значениями: **большой**, крупный, значительный, огромный, многочисленный, громадный; переносными значениями: великий, знаменитый, выдающийся, признанный. Британский национальный корпус предлагает следующие лексические варианты: **big**, large, biggish, bulky, cumbersome, huge, great, massive, gigantic, colossal, giant, spacious, immense, vast, palatial, considerable, substantial; Longman Language Activator вместе с номинативным представляет и переносное значение: having a big effect – major, considerable, great, huge, enormous, immense, tremendous [11].

Семантическое значение прилагательного «big» проявляется в определенном контексте или в паре с существительными. Количество подобных моделей может варьироваться от широкой сочетаемости до суженного диапазона, когда существует ограничение до употребления только с одним словом. Как правило, данное прилагательное характеризуется свободной природой, то есть возможностью лексем образовывать максимальное количество комбинаций с другими лексемами, соответственно, самыми частотными являются коллокации атрибутивного типа. Словарная статья современного авторитетного словаря английского языка Oxford Advanced Learner's Dictionary [12] предлагает следующие значения прилагательного «big»:

LARGE 1, OLDER 2, IMPORTANT 3, AMBITIOUS 4, POPULAR 5, ENTHUSIASTIC 6, DOING STH A LOT 7, GENEROUS 8.

В силу своего широкого семантического объема прилагательное «big» способно давать характеристику любому субъекту или объекту окружающей действительности, что позволяет ему сочетаться практически с любым конкретным существительным в своем прямом значении «большой по размерам»: this shirt isn't **big enough**; и переносном – «важный, знаменитый, известный»: a host of **big names** will be on the critical list when the wards reopen next Saturday (табл. 1).

По данным, которые были предоставлены программой корпус-менеджер, выделено 15,331,668 словоупотреблений прилагательного «big». Приведем некоторые примеры из конкорданса

Таблица 1. Контексты употребления BIG

British National Corpus (BNC)		SEARCH	FREQUENCY	CONTEXT	ACCOUNT
FIND SAMPLE: 100 200 500 1000 PAGE: << < 1 / 244 > >>					
CLICK FOR MORE CONTEXT HELP SAVE TRANSLATE ANALYZE					
1	CH1	W_newsp_tabloid	Q	. He didn't put on an act.' He just listened... this big star, the famous comedian, who was ready to try to understand all the	
2	CH1	W_newsp_tabloid	Q	their 20 biggest selling cars, including Renault and Citroen. And the group is big in gas-guzzling America, where there are an incredible 125 million cars on the road	
3	CH1	W_newsp_tabloid	Q	Holby Hospital -- the setting for the controversial BBC series Casualty. A host of big names will be on the critical list when the wards reopen next Saturday. Howards	
4	CH1	W_newsp_tabloid	Q	battering and a chilling nuclear power accident. It will also tackle the way a big city hospital adapts to becoming a Trust under the Government's controversial scheme. 1	
5	CH1	W_newsp_tabloid	Q	historians. # K.W.S., at No. 10 with Rock Your Baby, missed a big party in their honour after being mobbed by hysterical fans. The band were about	
6	CH1	W_newsp_tabloid	Q	be disappointed. Somewhere the story has lost its soul in the translation to the big screen and one is left with the fairly black and white conclusion that white Afrikaners	
7	CH1	W_newsp_tabloid	Q	prominence. -- # Give greedy banks the boot # IT'S disgusting that the big banks plan to charge fees to people in credit. They got themselves in a	
8	CH1	W_newsp_tabloid	Q	with presenter Anthea Turner to celebrate his return to TV. But goodness knows what Big Ears will think when he hears Noddy speaking with the voice of actress Susan	
9	CH1	W_newsp_tabloid	Q	motorists toot once for Amen and twice for Hallelujah. # Mon Dieu! This big daddy is no hero # WHO is that luscious young girl posing on the beach	
10	CH1	W_newsp_tabloid	Q	355 # OLDE TIMES: In Chester # Just TWO more cornettos # AND A BIG DOLLOP OD CULTURE IN FLORENCE # GEORGE THAW / DAVID KERR # THE proud citizens of	
11	CH1	W_newsp_tabloid	Q	with cement and sand. Now there are three waterfalls and a fountain. The big one in the centre tumbles down into a pool where goldfish swim. Pigeons and	
12	CH1	W_newsp_tabloid	Q	're giving our winners REAL diamonds in our sparkling new contest, as well as BIG cash prizes. Whether it's a skinny sliver of a thing or a rock	
13	CH1	W_newsp_tabloid	Q	feel like a million dollars too. So, instead of splashing out for the big one, we've got a number of gorgeous jewels to give away. Here	

Структура любой словарной статьи содержит дефиницию, включающую формальную и семантическую характеристику слова, ряд употреблений, иллюстрирующих сочетаемость единицы, и демонстрирующих системные семантические и стилистические связи, включая указания на тематические, лексико-семантические и ассоциативные группы; иллюстративные примеры реального употребления данного варианта как в узком контексте словосочетания, так и в более широком контексте речевого произведения.

Согласно данным словаря Oxford Advanced Learner's Dictionary (2020), наиболее частотными словосочетаниями являются: *big man, big house, big car, big boy, big dog, big smile, big problem, big surprise, big question, big difference*, где реализует прямое номинативное значение – большой по размеру, а также производное – сложный, важный, амбиционный.

Точно так же, как узкий контекст словосочетания обуславливает конкретное значение многозначного слова, в определенных условиях он может создавать семантическую диффузность, т.е. совместимость отдельных лексических значений, когда их разграничение не осуществляется, так как не представляется необходимым: *big star, big city, big party, big screen, big hats, big knife* [8].

Некоторые значения проявляются только в сочетании с определяющим словом, например, «*big deal*»: *...five paint brushes and 3.900 for a bathroom scales. Big deal!*; в некоторых сочетаниях значение полисемантической единицы представлено как фразеологически связанное, например «*Big Ears*» (*a man with unusual ears / Earesdropping on a conversation in which you do not participate.*): *but goodness knows what Big Ears will think when he hears Noddy speaking with the voice of actress Susan* [8].

Таблица 2. Частые коллокации **BIG** в двух регистрах корпуса. Правые коллокации

British National Corpus (BNC)							[HELP...]						
SEARCH		FREQUENCY			CONTEXT		ACCOUNT						
SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION)													
SEC 1 (ACADEMIC): 15,331,668 WORDS							SEC 2 (FICTION): 15,909,312 WORDS						
	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	ET-1	8	0	0.5	0.0	52.2	1	BLACK	62	0	3.9	0.0	389.7
2	CRIME	7	0	0.5	0.0	45.7	2	FAT	34	0	2.1	0.0	213.7
3	CITIES	20	8	1.3	0.5	2.6	3	MISTAKE	32	0	2.0	0.0	201.1
4	'	15	8	1.0	0.5	1.9	4	BROWN	30	0	1.9	0.0	188.6
5	BUSINESS	29	19	1.9	1.2	1.6	5	OLD	29	0	1.8	0.0	182.3
6	IS	14	10	0.9	0.6	1.5	6	ROOM	27	0	1.7	0.0	169.7
7	DIFFERENCE	7	9	0.5	0.6	0.8	7	FISH	26	0	1.6	0.0	163.4
8	PART	5	7	0.3	0.4	0.7	8	RED	24	0	1.5	0.0	150.9
9	BANG	12	33	0.8	2.1	0.4	9	SMILE	22	0	1.4	0.0	138.3
10	A	11	33	0.7	2.1	0.3	10	BLUE	21	0	1.3	0.0	132.0
11	BROTHER	6	26	0.4	1.6	0.2	11	TABLE	20	0	1.3	0.0	125.7
12	TO	5	25	0.3	1.6	0.2	12	YELLOW	20	0	1.3	0.0	125.7
13	CITY	5	29	0.3	1.8	0.2	13	HANDS	19	0	1.2	0.0	119.4
14	ENOUGH	17	131	1.1	8.2	0.1	14	ALLEN	18	0	1.1	0.0	113.1
15	AND	12	114	0.8	7.2	0.1	15	BED	18	0	1.1	0.0	113.1
16	FOR	7	74	0.5	4.7	0.1	16	FEET	18	0	1.1	0.0	113.1
17	ONE	5	66	0.3	4.1	0.1	17	DAY	17	0	1.1	0.0	106.9
18	.	7	108	0.5	6.8	0.1	18	DOUBLE	17	0	1.1	0.0	106.9
19	,	21	340	1.4	21.4	0.1	19	GREY	17	0	1.1	0.0	106.9
20	AS	6	110	0.4	6.9	0.1	20	WHEEL	17	0	1.1	0.0	106.9
							21	GREEN	16	0	1.0	0.0	100.6
							22	KITCHEN	16	0	1.0	0.0	100.6
							23	WOMAN	16	0	1.0	0.0	100.6

Особого внимания заслуживают экстралингвистические причины семантических изменений. Это, прежде всего, существование различного рода запретов или табу, продиктованных чувством деликатности, стремлением соблюдать приличия и языковой такт, когда речь идет, в частности, о физических недостатках. Например, использование эвфемизма «*big*» вместо оскорбительного «*fat*», в значении «толстый/ая», несколько смягчает негативные отношения к внешности человека, особенно если это касается женщины. В этой функционально-речевой ситуации прилагательное «*big*» приобретает, следуя тенденции политической корректности, определенную частность употребления, что очевидно

из данных корпуса (таблица 2, строка 23). Сочетание «big woman» зафиксировано в Британском национальном корпусе в разделе «правые коллокации», но не представлено ни в одной из проанализированных словарных статей.

Полисемантическое прилагательное «big» особенно интересно с точки зрения представленных в его структуре атрибутивных словосочетаний, где реализуется фразеологически связанные значения слова:

big chief – (informal) the person in charge of a business or other organization

big mouth – (informal) a person who talks a lot, especially about him- or herself, and who cannot keep secrets ► *big-mouthed* adj.

big-hearted – very kind; generous

big hitter – (informal) a person who is successful and has a lot of influence: they've appointed one of the industry's big hitters to the board.

a big cheese – (informal, humorous) an important and powerful person, especially in an organization

big fish – If you describe someone as a big fish, you believe that they are powerful or important in some way. [INFORMAL] The four men arrested were described as really big fish by the U.S. Drug Enforcement Agency. 2 phrase If you say that someone is a big fish in a small pond, you mean that they are powerful or important but only within a small group of people. [INFORMAL] In South Africa, Jani was a big fish in a small pond.

big man on campus – a successful popular male student at a college or university [9].

Все эти значения являются узуальными и зафиксированными во всех словарях современного английского языка. Однако, принимая во внимание данные Британского национального корпуса, следует особо отметить тот факт, что коллокация «**big fish**» приобретает высокую частотность употребления, что очевидно из данных корпуса (таблица 3, строка 6).

В целом полисемантическое прилагательное «big» сохраняет свою широкозначную природу и основное номинативное значение как в современных словарях, так и в лингвистическом корпусе данных остается неизменным – «большой по размеру, крупный». Номинативно-производные варианты данной лексической единицы остаются относительно статичными, реализуя значения «важный, амбиционный, популярный», однако, следуя данным корпуса, необходимо отметить коллокацию, которая находится на второй строке (правый коллокат) «**crime**», и, соответственно, приобретает значение «serious, atrocious, appalling, horrifying», вариант, который не нашел еще своего отражения в лексикографических источниках.

Таким образом, благодаря современным достижениям в области информационных технологий, изучение языковых процессов намного упрощается, соответственно, такое направление в сфере лингвистических исследований, как корпусная лингвистика предоставляет самые современные, легкодоступные и эффективные способы решения проблемы полисемии языковых единиц. Выбор определенной модели высказывания и соответствующих лексических единиц строится на синтагматическом принципе или линейных связях слов и является наиболее значимым в построении речи, подчиняясь определенным коммуникативным задачам. Изучение проблемы комбинаторной семантики на основе лингвистического корпуса данных следует считать актуальным и перспективным направлением исследования как системы словарного состава языка в целом, так и для выявления динамических процессов в сфере семантики в частности.

СПИСОК ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Архипова Н.Г. Сочетаемость слова в лексикографическом описании: дис. ... канд. филол. наук. М., 2000. 231 с.
2. Влавацкая М.В. Теоретические проблемы комбинаторной лингвистики. Лексикология. Лексикография. Новосибирск: Изд-во НГТУ, 2011. 508 с.
3. Гак В.Г. К проблеме семантической синтагматики, М.: АН СССР, Институт русского языка, 1972. 319 с. URL: https://www.phantastike.com/ru/yazykovyye_preobrazovaniya/pdf/
4. Захаров В.П. Корпусная лингвистика: учебно-методическое пособие: Изд-во СПбГУ, 2005. 48 с.
5. Подлеская В.И. Современные компьютерные методы в изучении и преподавании лингвистических дисциплин: корпусная лингвистика, М: РГГУ, 2011. URL: <http://www.rsuh.ru/print.html>
6. Рыков В.В. Прагматически ориентированный корпус текстов // Диалог-99. Москва–Таруса, 1999. 98 с.
7. Шмелёв Д.Н. Проблемы семантического анализа лексики. М.: Наука, 1973. 280 с.
8. British National Corpus. URL: <https://www.english-corpora.org/bnc/>
9. Collins W. Collins Cobuild Advanced Learner's Dictionary, HarperCollinsPublishers. 9th Edition, 2018. 6292 p.

10. Crystal David. A Dictionary of Linguistics and Phonetics / 6th ed., 2008. 528 p.
11. Longman Language Activator, Gate Edinburgh, Harlow, Essex CM20 2JE England and associated companies throughout the world, White Plains, N.Y.: Longman, 2nd edition 2022. 1530 p.
12. Oxford Advanced Learner's Dictionary, Oxford University Press 8th education 2010. 1796 p.; 10th education, 2020. 1820 p.
13. Sinclair J. The Idiom Principle Revisited // URL:gwern.net/doc/psychology/2014-chanturia.pdf

Поступила в редакцию 05.02.2023

Буйнова Ольга Юрьевна, кандидат филологических наук, доцент кафедры теории языка, межкультурной коммуникации и зарубежной литературы
ФГБОУ ВО «Удмуртский государственный университет»
426034, Россия, г. Ижевск, ул. Университетская, 1
E-mail: buinova@bk.ru

О.Ю. Буйнова

CORPUS-BASED APPROACH TO THE STUDY OF POLYSEMY

DOI: 10.35634/2412-9534-2023-33-4-790-797

The paper gives an insight into a semantic capacity of the adjective «big» with a corpus-based approach where British National Corpus helps to reveal the scope of meanings of this frequent lexical unit both in statics and dynamics; special attention is paid to combinatorial linguistics while analyzing functional speech aspects of the language on the whole and polysemy in particular. The relevance of the topic is connected with the fact that in all spheres of human activity, in linguistics as well, high technologies help to solve traditional problems of linguistics in a new and more effective way, their solution is achieved easier and more evidently. Accordingly, the purpose of the study is to identify how corpus linguistics helps to identify the ambiguity of words and systematizes the semantic structure of lexical units. The paper presents the key points concerning the processing of a large array of data and a contextual method to solve the problem of polysemy of words; a semantic analysis of a specific lexical unit is performed, taking into account the types of lexico-semantic variants and the method of statistical calculation. A comparative analysis of dictionary entries, which are presented both in electronic as well as in printed publications, of such authoritative lexicographic sources as: Collins Cobuild Advanced Learner's Dictionary, Longman Language Activator, Dictionary of Linguistics and Phonetics, Oxford Advanced Learner's Dictionary is carried out. As a result, the degree of semanticization of the lexical unit involved in the study is determined within the framework of micro- and macro-contexts.

Keywords: corpus linguistics, combinatorial lexicology, polysemy, lexico-semantic combinability, broadmeaningness of the adjective.

REFERENCES

1. Arkhipova N.G. Sochetaemost slova v leksikograficheskom opisanii: dis. kand. filol. nauk. [Word compatibility in lexicographic description: dis. candidate of Philology]. M., 2000, 231 p. (In Russian).
2. Vlavatskaya M.V. Teoreticheskie problemy kombinatornoj lingvistiki. [Theoretical Problems of Combinatorial Linguistics. Lexicology. Lexicography]. Novosibirsk: NSTU publisher. 2011, 508 p. (In Russian).
3. Gak V.G. K probleme semanticheskoy sintagmatiki [On the problem of semantic syntagmatics]. Moscow: USSR Academy of Sciences, Institute of the Russian Language, 1972, 319 p. (In Russian).
4. Zakharov V.P. Korpusnaya lingvistika: uchebno-metodicheskoe posobie. [Corpus linguistics: an educational and methodological manual]. St. Petersburg: Publishing House of St. Petersburg State University, 2005, 48 p. (In Russian).
5. Podlesskaya V.I. Sovremennye kompyuternye metody v izuchenii i prepodavanii lingvisticheskikh disciplin: korpusnaya lingvistika. [Modern computer methods in the study and teaching of linguistic disciplines: corpus linguistics]. Moscow: RTSU, 2011. // URL: <http://www.rsuh.ru/print.html> (In Russian).
6. Rykov V.V. Pragmaticheski orientirovannyj korpus tekstov. [Pragmatically oriented corpus of texts] // Dialog-99 - Moscow-Tarusa, 1999. 98 p. (In Russian).
7. Shmelev D.N. Problemy semanticheskogo analiza leksiki. [Problems of semantic analysis of vocabulary]. M: Nauka Publishing House, 1973, 280 p. (In Russian).
8. British National Corpus // URL: <https://www.english-corpora.org/bnc/> (In English).
9. Collins W. Collins Cobuild Advanced Learner's Dictionary, HarperCollinsPublishers 9th Edition 2018, 6292 p. (In English).
10. Crystal David A Dictionary of Linguistics and Phonetics / David Crystal. – 6th ed., 2008, 528 – 117 p. (In English).

11. Longman Language Activator, Gate Edinburgh, Harlow, Essex CM20 2JE England and associated companies throughout the world, White Plains, N.Y.: Longman, 2nd edition 2022, 1530 p. (In English).
12. Oxford Advanced Learner's Dictionary, Oxford University Press 8th edition 2010, 1796 p.; 10th edition, 2020, 1820 p. (In English).
13. Sinclair J. The Idiom Principle Revisited // URL:gwern.net/doc/psychology/2014-chanturia.pdf (In English).

Received 05.02.2023

Buinova O.Yu., Candidate of Philology, Associate Professor at Department of Theory of Language,
Intercultural Communication and Foreign Literature
Udmurt State University
Universitetskaya st., 1, Izhevsk, Russia, 426034
E-mail: buinova@bk.ru