

**КОРПУСНЫЕ ДАННЫЕ В РАЗРАБОТКЕ
УЧЕБНЫХ СЛОВАРЕЙ СОЧЕТАЕМОСТИ**

Гончаренко И. Г.

Российский государственный профессионально-педагогический университет
(Екатеринбург, Россия)ORCID ID: <https://orcid.org/0000-0001-5515-9332>

Аннотация. В статье рассматривается проблема использования корпусных данных в учебной лексикографии. Данные традиционных учебных словарей сочетаемости, таких как «Оксфордский словарь сочетаний для изучающих английский язык», сопоставляются с данными Британского национального корпуса, насчитывающего около ста миллионов словоупотреблений британского варианта английского языка в его письменной и устной форме и считающегося одним из наиболее сбалансированных корпусов (содержащим тексты различных жанров и сфер функционирования). Обращение к корпусу осуществлялось с помощью корпусного менеджера (набора доступных онлайн-инструментов для поиска и извлечения данных), благодаря использованию которого возможно анализировать данные, создавая статистические запросы и запросы ключевых слов, конкордансы (списки всех случаев употребления запрашиваемого языкового выражения в контексте), списки сочетаний – коллокаций и т. п.

Кроме того, в статье сопоставляются данные «Словаря словосочетаний» французского языка с данными корпусно-ориентированного электронного словаря «Антидот» канадской компании «Друид». Данный словарь представляет собой множество словарей, объединенных одной программной оболочкой. Каждое слово в словаре помимо произношения, его грамматических форм, этимологии и т. п. сопровождается индексом частотности, который указывает на относительную частотность слова в корпусе из шести миллиардов слов, использовавшемся при разработке словаря. Наиболее ценной частью данной программы является наличие словаря сочетаний, который содержит все наиболее значимые сочетания заголовочного слова с другими словами (главными для него или зависимыми), сгруппированные по признакам синтаксической роли в предложении и частотности.

Новизна работы заключается в демонстрации возможностей использования корпусных данных в учебной лексикографии (в частности, при создании учебных словарей сочетаемости). На конкретных примерах показано, как лингвистический корпус позволяет значительно расширить представление о семантических, стилистических и синтаксических особенностях слов. Сделан вывод о преимуществах использования корпусных данных перед использованием традиционных словарей, отмечены ограничения корпусных данных при синтаксическом и семантическом анализе. В заключение предлагается проект разработки основанного на корпусных данных обучающего словаря для изучающих русский язык как иностранный.

Ключевые слова: корпусные исследования; Национальный корпус русского языка; Британский национальный корпус; Antidote Druide; лексическая сочетаемость; словарь сочетаемости; атрибутивные сочетания

Для цитирования: Гончаренко, И. Г. Корпусные данные в разработке учебных словарей сочетаемости / И. Г. Гончаренко. – Текст : непосредственный // Филологический класс. – 2023. – Т. 28, № 2. – С. 55–68.

**USING LANGUAGE CORPORA IN BUILDING
LEARNER'S COMBINATORIAL DICTIONARIES**

Iliia G. Goncharenko

Russian State Vocational Pedagogical University (Ekaterinburg, Russia)

ORCID ID: <https://orcid.org/0000-0001-5515-9332>

Abstract. The article deals with the problems of using corpus data in educational lexicography. The data from traditional collocations dictionaries, such as the *Oxford Collocations Dictionary for Students of English*, are compared with the data extracted from the British National Corpus (BNC). The BNC is an approximately 100-million-word corpus of written and spoken British English (it is considered as a balanced corpus that contains texts from a wide range of different language genres and text domains). A corpus manager (a web-based tool for searching and retrieving lexical, grammatical and textual data) was employed in the study. Due to this it has become possible to analyze the data, generating frequency information, concordances (i.e. lists of all of the occurrences of a particular search term in a corpus, presented within the context in which they occur), keywords, collocations or carrying out statistical tests.

In addition, the data from *Dictionnaire des combinaisons de mots* are compared with the data from the corpus-based electronic dictionary *Antidote of the Canadian software company Druid informatique*. This program comprises multiple dictionaries placed within a unified interface. The entry for each word displays its pronunciation, inflected forms, etymology, etc. along with their respective frequency. A frequency index is provided for each word; it indicates the relative frequency of the word in the six billion-word corpus. The presence of a dictionary of collocations that provides all the most significant combinations of the entry word with other words (functioning either as leading or dependent components), grouped by their syntactic function in the sentence and frequency is the most valuable feature of this program.

The novelty of the work lies in the fact that it demonstrates the educational potential of corpus data in lexicography, in particular, in the field of compiling collocation dictionaries. The specific examples show how linguistic corpora can help comprehend the semantic, stylistic and syntactic specific features of words. The paper concludes that corpus data has many advantages over traditional dictionaries; at the same time, the limitations of corpus data in syntactic and semantic analysis are noted. In conclusion, the authors outline a project for developing a corpus-based pedagogical dictionary for students of Russian.

Keywords: corpus study; Russian National Corpus (RNC); British National Corpus (BNC); Antidote Druide; lexical collocability; collocations dictionary; attributive collocations

For citation: Goncharenko, I. G. (2023). Using Language Corpora in Building Learner's Combinatorial Dictionaries. In *Philological Class*. Vol. 28. No. 2, pp. 55–68.

Статус словаря в эпоху корпусных данных

Произошедшее за последние десятилетия проникновение информационных технологий в различные виды человеческой деятельности существенным образом повлияло на них. Фундаментальные трансформации произошли и в лингвистической науке. Применение информационных технологий привело к возникновению компьютерной лингвистики, к числу наиболее актуальных направлений которой можно отнести корпусную лингвистику.

Согласно одному из определений «под лингвистическим, или языковым, корпусом текстов (или обычно просто *корпусом текстов*) понимается большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» (курсив авторский – И. Г.) [Захаров 2020: 11]. Если исключить из определения наличие «филологической» и «лингвистической» составляющих, то под корпусом можно понимать все проиндексированные поисковой машиной

тексты, написанные на естественном или искусственном языках, – глобальная сеть будет рассматриваться в таком случае как банк данных, а поисковая машина – как корпусный менеджер.

Идея рассмотрения веба как корпуса, разумеется, не нова [Kilgarriff, Grefenstette 2003], и в настоящее время она успешно реализуется при создании монографических исследований (см., например: [Русский язык коронавирусной эпохи 2021]), составлении словарей – частотных [Sharoff, Umanskaya, Wilson 2013], толковых, – в особенности тех, которые посвящены описанию актуального состояния языка [Леонтьева, Щетинина 2021; Словарь русского языка коронавирусной эпохи 2021]; очевидно, что подобные издания были бы невозможны без использования корпусных данных и обращения к вебу как корпусу. Корпусные менеджеры (а в случае использования веба как корпуса поисковые машины) обладают довольно гибкими системами формирования пользовательских запросов, которые позволяют получать данные о грамматических, сочетаемых, стилистических особенностях запрашиваемых языковых выражений.

Ни в коем случае не абсолютизируя использование корпусных данных, однако принимая во внимание постоянное увеличение их объема, а также совершенствование технологий обработки естественного языка, можем указать на намечающуюся тенденцию к постепенной утрате традиционными лексикографическими изданиями их самых важных функций – информативной и нормативной. Обладая в сравнении с вебом рядом существенных недостатков (к которым следует отнести неизбежное отставание от развития языка, ограниченность объема, скорость поиска, цену), в обозримой перспективе ориентированные на массовую аудиторию лингвистические словари могут перестать играть роль посредников между языком и теми, кто этим языком как средством коммуникации пользуется.

Сказанное заставляет задаться вопросами о том, какими должны быть словари в информационную эпоху, что нового по сравнению с корпусами (и вебом как корпусом) они могут предложить, как в них могут быть использованы большие данные, касающиеся употребления языка, какие сложности могут возникнуть при составлении новых словарей.

Словари активного типа: возможности использования при изучении русского языка как иностранного

В связи с переключением внимания современной лингвистики с языка как замкнутой в себе системы на язык как средство коммуникации и мышления получают распространение новые интерпретации ставших классическими лингвистических идей – теории функционального синтаксиса, коммуникативной грамматики, грамматики конструкций и т. п.

Тенденция к исследованию особенностей языка, делающих его средством хранения и передачи информации, проявляется и в лексикографической практике.

Не говоря о всевозможных ассоциативных словарях, в том числе учебных, ориентированных на изучающих русский язык как иностранный (РКИ) [Тарасов 2017], отметим, что даже созданный в традициях классической отечественной лексикографии «Русский семантический словарь» с реализованным в нем идеографическим принципом описания

лексики призван открыть «перед читателями исторически сложившуюся в языке, им самим созданную картину мира» [Русский семантический словарь 2002: xvi]. Таким образом обозначается цель одновременно и лингвистическая (описание лексической системы русского языка) и лингвокультурологическая (описание заключенной в языке «картины мира»).

Наиболее интересным примером применения нетривиальных подходов (а также использования корпусных данных) при описании лексики является «Активный словарь русского языка» (АС). Как следует уже из названия, это словарь активного типа, т. е. такой, который должен «обеспечить нужды говорения, или, более широко, нужды производства текстов» [Перспектив активнаго словаря русскогo языка 2010: 18]. Весьма примечательно, что среди предшествующих, сходных с данным словарем по задачам и идеологии словарей указываются словари учебные, ориентированные на изучающих язык как иностранный, такие как «Longman Dictionary of Contemporary English» (LDCE) или «The Advanced Learner's Dictionary of Current English» (сейчас известный как «Oxford Advanced Learner's Dictionary» (OALD)).

Опуская описание структуры словарной статьи и стоящей за ней методологии, отметим, что и толкование лексических значений с использованием особого метаязыка, и представление всех возможных моделей управления лексической единицы, и большое количество примеров сочетаний, в которых она может выступать, и характеристика синонимических и антонимических отношений, в которые она вступает, и характеристика особенностей просодии делают АС чрезвычайно ценным для изучающих русский язык как иностранный источником информации.

Как известно, для эффективного – активного – овладения лексической системой неродного языка важно не только располагать сведениями о значении составляющих ее лексических единиц, их грамматических и синтаксических признаках, стилистической окраске и т. п., но и представлять комбинаторные возможности каждой лексемы. В прямом предшественнике АС – «Толково-комбинаторном словаре русского языка» (ТКС) – эти возможности назывались лексическими

функциями [Толково-комбинаторный словарь русского языка 2016: 94].

При всей той ценности, какую имеют ТКС и АС для изучающих РКИ (и ценность эта прежде всего определяется широтой описания комбинаторных возможностей лексем – их синтагматики, а также многообразием представления для них синонимов, аналогов, конверсивов, антонимов, дериватов – парадигматики), данные издания сложно назвать учебными. Разумеется, они могут использоваться для решения учебных задач, однако в большей степени они ориентированы на академические, исследовательские цели – и в качестве учебных едва ли будут использоваться повсеместно. Отметим также и то обстоятельство, что в ТКС описано около 286 лексических единиц, а последним опубликованным томом АС на настоящий момент является том третий, завершающийся словами на букву «З». Таким образом, использовать как законченные и самодостаточные лексикографические произведения данные словари пока нельзя.

Кроме ТКС, словник которого весьма ограничен, и АС, работа над которым еще не завершена, информацию о комбинаторных особенностях слов русского языка можно получить в «Словаре сочетаемости слов русского языка» [Словарь сочетаемости слов русского языка 1983]. Не подвергая сомнению большую методическую ценность словарей подобного типа, отметим, что такие словари в большей степени являются справочными, нежели учебными, представленный здесь материал не имеет связи с уровнем владения русским языком потенциальным читателем, не учтенной является и частотность конструкций, в то время как и для осваивающих иностранный язык, и для преподающих его признак частотности, причем частотности, распределенной по сферам функционирования (разговорная речь, научный стиль, публицистика и т. п.), является решающим при выборе языкового материала.

Приходится констатировать, что в отечественной лексикографии пробел, связанный с отсутствием словарей активного типа, подобных Лонгмановским или Оксфордским,

на который сетовали авторы «Перспекта активного словаря русского языка», до сих пор остается незаполненным. Как заполнить этот пробел и есть ли необходимость в его заполнении вообще – на эти вопросы попытаемся ответить далее.

Словарь сочетаемости vs корпус

В лексикографической практике, посвященной описанию английского языка, справедливо считающейся одной из самых развитых во всей европейской лексикографии, разработаны не только приемы составления учебных словарей, посвященных всестороннему и в то же время сбалансированному и методически хорошо продуманному описанию лексики, но и приемы создания словарей сочетаемости – таких как «The BBI Dictionary of English Word Combinations» [Benson 2010], «Macmillan Collocations Dictionary for Learners of English» [Macmillan Collocations Dictionary 2010], «Longman Collocations Dictionary and Thesaurus» [Longman Collocations Dictionary and Thesaurus 2013]. Словари сочетаемости ориентированы на то, чтобы позволить изучающему язык перейти от пассивного понимания лексического значения к активному использованию слова в речи. С точки зрения возможностей «активации» пассивного словаря рассмотрим один из популярных и авторитетных словарей сочетаемости «Oxford Collocations Dictionary for Students of English» (OCD), сопоставив его с другим доступным средством «активации» словаря – Британским национальным корпусом¹ (данный корпус выбран по причине его сбалансированности, а корпусный менеджер – по причине гибкости создаваемых с его помощью запросов).

Сочетания (здесь и далее в связи с сопоставлением данных OCD и BNC как эквивалент термина «collocation» будем использовать выражения «сочетание» / «сочетание слов», не вдаваясь в различия между терминами «коллокация», «коллигация», «конструкция» и т. п., см. об этих различиях в отечественной литературе: [Влавацкая 2016: 75–93]) в словаре OCD реализуют модели прилагательное + существительное, глагол + существительное, существитель-

¹ British National Corpus. URL: <https://www.english-corpora.org/bnc/> (mode of access: 25.08.2022).

ное + глагол и т. п.; кроме того, в словарь включены короткие идиоматичные выражения.

Такое описание позволяет охарактеризовать не только особенности лексической сочетаемости слов, но и их способности управлять зависимыми компонентами и выполнять различные синтаксические функции. При этом важно, что в словаре представлено 9000 заголовочных слов, лексическое значение которых уже должно быть известно читателю, владеющему английским языком на уровне Upper Intermediate.

В стилистическом плане в словаре представлен прежде всего книжный язык, который составители называют «moderately formal language» [Oxford Collocations Dictionary 2002: viii]. Как и OALD и LDCE, это словарь активного типа, ориентированный на речепорождение.

Как все современные авторитетные учебные словари английского языка, OCD составлен с учетом корпусных данных (об использовании корпусов в лексикографических целях см., в частности: [Kilgarriff 2022; Kosem 2016]), а именно данных BNC.

Однако, на наш взгляд, этот учебный словарь не способен дать необходимой для изучающего язык картины сочетаемости лексических единиц: здесь нет информации о том, в каком из значений чаще используется слово (соответственно какие сочетания с ним встречаются чаще), какие из образованных с ним сочетаний предпочтительнее и в каких контекстах (как уже отмечалось, с точки зрения методики эта информация представляется чрезвычайно важной), – и именно эту информацию можно извлечь из корпусных данных.

В качестве примера, иллюстрирующего преимущества использования корпусных данных перед использованием комбинаторных словарей, сопоставим данные BNC со сведениями из OCD о сочетаемости близких по лексическому значению и по этой причине способных вызвать у изучающих английский язык (в особенности, если в их родном языке нет эквивалентной синонимии) сложности в семантической дифференциации лексем *habit* (привычка) и *custom* (обычай) с атрибутивными словами.

В «Новом большом англо-русском словаре» дефиниция первых значений лексем выглядит почти одинаково: *custom* – «п 1. обычай; обыкновение, привычка» [Новый большой англо-русский словарь, т. 1: 499]; *habit* – «п 1. привычка; обычай; обыкновение» [Новый большой англо-русский словарь, т. 2: 91]. Сходство значений усиливается указанием на то, что обе лексемы могут входить в состав вариантов одного и того же идиоматического выражения *it is a custom / habit with sb to do sth* (примеры: *it was a custom with him to get up early; it is a habit with him to keep early hours*). Ср. указание на синонимии в «The BBI Combinatory Dictionary of English»: *habit* определяется через *custom* – «**habit** n. [“custom”] [“usual manner”]» [Benson 2010: 180].

В OCD с *habit* указаны следующие атрибутивные сочетания:

habit noun

• ADJ. **annoying, anti-social, bad, dirty, disconcerting, horrible, irritating, nasty, unfortunate** *Life has a nasty habit of repeating itself.* | **charming** (often ironic), **endearing, good** *one of his more endearing habits* ◊ *her charming habit of setting fire to cats* | **eccentric, odd | old | daily, regular | personal, sexual, social** *I found some of his personal habits rather disconcerting.* | **buying, shopping, spending** *an effort to change the buying habits of the British public* | **dietary, drinking, eating, feeding** | **reading, viewing** *women's television viewing habits* | **drug, smoking** *trying to kick the smoking habit* [Oxford Collocations Dictionary 2002: 360].

Здесь представлена довольно широкая сочетаемость, атрибутивные слова сгруппированы в соответствии с их семантикой, приведены примеры контекстов употребления некоторых из возможных сочетаний. Но если сопоставить эти сведения с той частью посвященной существительному *custom* словарной статьи, в которой приведены только атрибутивные сочетания, то можно получить весьма скудную информацию:

custom noun

• ADJ. **accepted, age-old, ancient, established, old, traditional | quaint | local | social** [Oxford Collocations Dictionary 2002: 185].

Здесь есть два пересечения сочетаний (прилагательные *old* и *social* сочетаются как с *habit*, так и с *custom*), но нет примеров контекстов употребления сочетаний кроме

того, те атрибутивные слова, которые здесь представлены, могут тоже сочетаться с *habit*.

Если же сравнить частотность препозитивных атрибутивных сочетаний именных

лексем *habit* и *custom* (с минимальным, т. е. равным единице, расстоянием между элементами) в BNC (запрос вида: HABIT_nn / CUSTOM_nn + ADJ)¹, можно получить картину, представленную в таблице 1.

Таблица 1. Количество вхождений сочетаний «АТРИБУТИВНОЕ СЛОВО + CUSTOM / НАВИТ» в BNC

		вхождений			вхождений	
		habit	custom		custom	habit
1	BAD	139	3	LOCAL	54	2
2	OLD	63	30	SOCIAL	38	24
3	SMOKING	54	0	OLD	30	63
4	GOOD	26	2	ANCIENT	19	0
5	NASTY	25	0	FRENCH	15	3
6	DIETARY	25	1	NEW	12	11
7	SOCIAL	24	38	ENGLISH	11	5
8	SEXUAL	23	1	ESTABLISHED	10	2
9	PERSONAL	20	0	TRADITIONAL	10	3
10	REGULAR	14	2	QUAINT	9	0

В таблице приведены первые 10 записей, которые упорядочены по убыванию вхождений сочетаний, заливкой отмечены пересечения вхождений по атрибутивным словам.

В ходе сопоставления данных BNC с OCD можно увидеть, что все представленные в словарных статьях OCD атрибутивные слова есть и здесь (отсутствие некоторых прилагательных в таблице объясняется тем, что они не вошли в первые 10 записей), но они даны в сопоставлении и в контекстах. Всего для *custom* получено 278 атрибутивных препозитивных сочетаний, для *habit* 439 (отметим, что и эта статистическая информация обладает определенной ценностью – методы оценки статистической релевантности результатов запросов коллокаций описаны в работах: [Хохлова: 2010; Kormacheva, Pivovarova, Korotev 2018; Stefanowitsch 2020: 217 et seq.]).

Беглый взгляд на прилагательные, сочетающиеся с *custom*, позволяет выявить большое количество определителей, основным значением которых является указание на принадлежность к какой-либо

национальности (*Arab, British, Egyptian, French, Greek, Irish, Italian, Jewish, Korean, Spanish* и др.), это целый класс объединенных общей семантикой прилагательных, который не представлен в словарной статье в OCD. Между тем конструкция «*custom* + атрибутивное слово с семантикой принадлежности к какой-либо общности» имеет очень широкое распространение, о чем свидетельствует то, что 50 прилагательных из полученных в результате запроса 278 содержат указание на национальность; помимо этого есть указатели на религиозную (*Christian, pagan*), историко-культурную (*medieval, barbarian*), социально-организационную (*tribal*), странственную (*local, regional, western*) и др. типы общности. В целом около трети всех атрибутивных слов, сочетающихся с *custom*, по данным BNC, имеют значение, связанное с какой-либо социальной общностью.

Очевидно также, что *habit* значительно чаще, чем *custom*, вступает в сочетания с глагольными образованиями. Результаты запроса сравнительной сочетаемости *habit* и *custom* с глагольными

¹Адрес запроса: <https://www.english-corpora.org/bnc/?c=bnc&q=108315590> (mode of access: 25.08.2022).

ми образованиями на -ing (запрос вида HABIT_nn / CUSTOM_nn + _v?g)¹ представлены в таблице 2 (выборка ограничена числом вхождений, равным 3).

Таблица 2. Количество вхождений сочетаний «ГЛАГОЛЬНОЕ ОБРАЗОВАНИЕ НА -ING + CUSTOM / HABIT» в BNC

№	Атрибутивное слово	Число вхождений		Атрибутивное слово	Число вхождений	
		Habit	Custom		custom	Habit
1	DRINKING	7	0	PAYING	3	0
2	READING	7	0	PROVIDING	2	0
3	BUYING	5	0	REMOVING	2	0
4	FEEDING	5	0	LOSING	1	0
5	SMOKING	5	0	ILLUSTRATING	1	0
6	PRESCRIBING	3	0	GETTING	1	0
7	PURCHASING	3	0	FOLLOWING	1	0

Любопытно, что значительную часть приведенных форм составляют глагольные образования со значением приобретения, потребления, присвоения и т. п., в то время как сочетаемость *custom* с этими словоформами весьма ограничена. Следует отметить также более

высокую частотность форм множественного числа существительного *habit* в сочетаниях с глагольными формами. Так, при введении запроса вида **ing + HABIT*² получим картину, представленную в таблице 3 (выборка ограничена числом вхождений, равным 10).

Таблица 3. Количество вхождений сочетаний «ГЛАГОЛЬНОЕ ОБРАЗОВАНИЕ НА -ING + HABIT» в BNC

№	Сочетание	Число вхождений	Графическое представление числа вхождений
1	EATING HABITS	135	
2	SMOKING HABITS	42	
3	DRINKING HABITS	24	
4	READING HABITS	21	
5	BUYING HABITS	19	
6	SMOKING HABIT	19	
7	FEEDING HABITS	18	
8	SHOPPING HABITS	11	
9	ANNOYING HABIT	10	
10	WORKING HABITS	10	

Не составляет особого труда заметить, что в ряду приведенных сочетаний преобладают сочетания с формами множественного числа существительного *habit*. Так, *eating habits* встречается 135 раз, в то время как как *eating habit* – 2; сходным является соотношение сочетаний *buying habits / habit* (19 с множественным числом, 2 с единственным), *reading habits / habit* (21 / 7), *viewing habits / habit* (8 / 1),

shopping habits / habit (11 / 1), *drinking habits / habit* (24 / 1). Очевидно, что здесь наблюдается реализация одной и той же устойчивой модели синтаксической конструкции, аналогом которой в русском языке будет структура «атрибутивное слово с семой узуального действия + привычки» (*пищевые привычки, потребительские привычки, читательские привычки, зрительские привычки и т. д.*).

¹ Адрес запроса: <https://www.english-corpora.org/bnc/?c=bnc&q=108319736> (mode of access: 25.08.2022).

² Адрес запроса: <https://www.english-corpora.org/bnc/?c=bnc&q=108385240> (mode of access: 25.08.2022).

Аналогичные запросы можно составить со словами других морфологических классов – в результате можно получить детальную картину сочетаемостных, семантических, идиоматических свойств лексем *habit* и *custom*.

Конечно, необходимо учитывать, что, хотя BNC считается наиболее хорошо разработанным и сбалансированным корпусом английского языка, наполнение BNC завершилось в 1994 году, поэтому для получения информации об актуальном употреблении запрашиваемых лексических единиц придется обращаться к другим корпусным данным. Кроме того, большинство современных корпусных менеджеров не способно обрабатывать синтаксические запросы (о проблематике синтаксической разметки на материале немецкого корпуса см., например: [Meurers, Müller 2009]), соответственно, при получении результатов запроса пользователь будет иметь дело с простыми совпадениями, а не с итогом синтаксического анализа, и в полученных в ходе обработки запроса данных неизбежно будут появляться нерелевантные, с точки зрения пользователя, записи. Так, в выдаче результатов запроса *ing + HABIT были записи, в которых существительное *habit* было прямым дополнением глагола, стоящего в одном из времен Continuous (примеры типа: *I'm forming habits and staking claims to little pieces of territory*), что формально соответствовало поисковому запросу. Еще больше примеров, в которых запрашиваемое слово будет зависимым от глагола элементом, ожидаемо демонстрирует запрос *ing + CUSTOM – и такое распределение синтаксических функций между *habit* и *custom* также будет весьма информативным.

Подобных записей может быть много, и их, безусловно, следует использовать осторожно, обращая внимание на случаи грамматической омонимии и учитывая синтаксические функции запрашиваемых элементов, но, несмотря на указанные сложности, возникающие при использовании корпусных менеджеров, информация, извлекаемая из корпусных данных, отличается большей репрезентативностью и полнотой по сравнению с информацией, содержащейся в словарях, – даже таких ориентированных на использование корпусных данных, как «Longman Collocations Dictionary and Thesaurus» [Longman

Collocations Dictionary and Thesaurus 2013]; с дидактической же точки зрения способ получения информации путем формирования запросов к корпусным данным представляется тем более эффективным, поскольку он предполагает активный поиск информации, а не пассивное ее получение, – и большую вовлеченность в поисковый процесс (в англоязычной лингводидактике такой тип изучения языка получил название «data-driven learning» [Baker 2006: 54], в отечественной лингводидактике такой подход тоже постепенно получает распространение [Лебедева 2020]).

Следует признать, что существующие лингвистические словари ориентированы прежде всего на представление лексики. Конечно, зачастую словарные статьи содержат примеры употреблений слов в словосочетаниях и предложениях, но эти примеры обычно являются именно «иллюстрациями» данных составителями дефиниций, а не контекстами, на основе анализа которых читатель сам бы мог сформировать представление о лексическом значении слова.

Известно, что лексическое значение всегда реализуется в каком-то контексте, поэтому при обучении иностранному языку важно предлагать наиболее типичные контексты употребления изучаемых слов. Однако для выработки навыка уверенного использования слова в речи минимального контекста, описываемого в терминах лексических функций (как в ТКС) или представленного в виде набора сочетаний и неэксплицируемого с помощью метаязыка (как в АС), на наш взгляд, недостаточно. Минимальной единицей речевой коммуникации является высказывание (предложение), и наилучшим способом представления лексемы в учебном словаре была бы демонстрация ее синтаксических функций.

Словарь сочетаемости vs словарь-корпус

Ряд недостатков «лексико-ориентированной» идеологии, реализуемой при составлении лингвистических словарей, отсутствует в некоторых современных электронных словарях, среди которых прежде всего следует назвать словарь Antidote канадской компании *Druide informatique*. В отечественной лингводидактической литературе данный словарь упоминается чрез-

вычайно редко, поэтому считаем необходимым дать его краткое описание.

Программа Antidote представляет собой множество словарей французского – а в последних версиях и английского – языков, объединенных одной программной оболочкой. Помимо толкового словаря с вполне традиционными дефинициями, в нем есть словарь синонимов, антонимов, представлены лексические поля, состоящие из групп слов, объединенных семантически, словообразовательные гнезда, таблицы спряжения глаголов, рифмы, цитаты, содержится этимологическая информация, а также справочная информация по орфографии и пунктуации, фонетике, грамматике, редактированию, деловому письму. Таким образом, этот словарь имеет ярко выраженный практически-нормативный характер.

По утверждениям разработчиков, в словарь версии Antidote 11 вошли 133 тысячи лексических единиц, среди которых 16000 имен собственных, 900 тысяч сочетаний, синонимические, гипер- и гипонимические связи установлены между лексическими единицами, количество которых насчитывает более чем 1000000, в словаре дано 50000 примечаний, касающихся трудностей французского языка, озвучено 1,6 миллиона словоформ с парижским и монреальским произношением (в английской части озвучено 500 тысяч словоформ с лондонским и торонтским произношением), в англо-французской и француско-английской частях содержится 2,7 миллиона переводов (переводятся не только заголовочные слова, но и словосочетания, в которые они входят). В исторической части словаря прослеживается более 3 миллионов этимологических связей, характеристика эволюции значений слов, изменения их фонетических и графических обликов сопровождается примерами употреблений начиная с XI века.

Но принципиальным отличием этого электронного словаря от традиционных словарей является представленное здесь разнообразие сочетаний лексических единиц, называемых совместными встречаемостями (*les cooccurrences*), которое стало возможным благодаря использованию корпусных данных. Материалом для словаря послужили данные франкофонной части интернета – сайтов

СМИ, цифровых библиотек и других электронных ресурсов. Как утверждают составители, корпус данных, использованных для словаря, насчитывает 6 миллиардов слов (270 миллионов фраз), при этом использованные для выявления сочетаемости методы анализа языкового материала позволили установить синтаксические связи между отдаленными (разделенными другими словами) компонентами предложений [Guide d'utilisation d'Antidote 11 2021: 83]. Сочетания сопровождаются индексом от 0 до 100, показывающим относительную частотность слова в корпусе.

В качестве примера, иллюстрирующего возможности словаря при анализе сочетаемости, рассмотрим сочетания с глаголом *lire* (*читать*). Сочетаемость глагола иллюстрируется конструкциями, в которых глагол употребляется: 1) с подлежащим, т. е. выступает в роли сказуемого (*le lecteur lit* – всего 47 сочетаний, каждое из которых может сопровождаться примерами из нескольких – обычно 5–7 – предложений), 2) с прямым дополнением (*lire l'article* – 251 сочетание), 3) с любым другим непрямым дополнением (*lire dans le journal* – 164 сочетания), 4) с обстоятельственным адвербиальным распространителем (*lire à haute voix* – 93 сочетания); 5) как определение (*lettre lue* – 48 сочетаний); 6) как распространитель имени существительного в номинативном словосочетании (*temps de lire* – 16 сочетаний); 7) как распространитель глагола (*savoir lire* – 52 сочетания); 8) как распространитель имени прилагательного (*agréable à lire* – 22 сочетания); 9) как компонент словосочетания с сочинительной связью (*lire et écrire* – 53 сочетания). Таким образом, весьма подробно оказывается описан репертуар сочетаемостных возможностей глагола *lire* (в совокупности представлено 746 сочетаний и несколько тысяч предложений), реализующего различные синтаксические функции.

Antidote позволяет получать информацию о сочетаемости лексических единиц в сопоставлении (функция *croiser avec un autre mot*), но, в отличие от результата обработки запроса о сопоставительной сочетаемости в BNC, здесь будут выдаваться только результаты совпадений (пересечений) сочетаний.

Любопытно сопоставление данных Antidote и классического словаря сочетаемо-

сти – «Dictionnaire des combinaisons de mots» (DCM) на примере атрибутивных сочетаний с существительными *coutume* и *habitude* [Dictionnaire des combinaisons de mots 2007: 217, 411–412]. В DCM атрибутивные слова и конструкции разделены на изолированные в семантическом отношении группы, объединенные указанием на социальную общность, место, время, оценку. В Antidote определители разделены на две группы – описательные прилагательные (*adjectifs descriptifs*) и классифицирующие прилагательные (*adjectifs classificateurs*) – такое деление является скорее формальным, однако при формировании запросов в корпусных менеджерах даже такое деление труднодостижимо.

В DCM в частях словарных статей, посвященных атрибутивным словосочетаниям с лексемами *habitude* и *coutume*, видим всего 3 пересечения (*habitude / coutume + bien ancrée; ancienne + habitude / coutume; vieille + habitude / coutume*), в то время как в словаре Antidote таких пересечений 14. С лексемой *habitude* в DCM представлено 18 атрибутивных словосочетаний (в Antidote – 85, при этом в Antidote входят 14 из 18 сочетаний DCM), с лексемой *coutume* в DCM – 37 словосочетаний (в Antidote – 55, при этом в него входят 19 из 37 DCM).

На основе количественного и качественного сопоставления данных двух словарей можно сделать вывод о том, что в Antidote ожидаемо представлен более широкий круг сочетаемостей, при этом каждое словосочетание, как правило, сопровождается несколькими примерами предложений; пересечения прилагательных наблюдаются со всеми рядами, указанными в словарных статьях в DCM (исключения составили только в определенной степени идиоматичные *belle coutume* и *petite habitude*). В обоих словарях имеем дело с указанием на употребление множественного числа *habitude*: в DCM на это указывают примеры (*les habitudes consuméristes ont pris le dessus*); в Antidote сами сочетания имеют форму множественного числа (*habitudes tabagiques, habitudes télévisuelles, habitudes gastronomiques*).

Оставляя за пределами рассмотрения очевидные преимущества любых электронных словарей и их веб-версий (сколь угодно большое количество языкового материала, быстрота поиска и удобство навигации,

многофункциональность и т. п.), многократно описанные в литературе [Воронцова 2022; Dziemianko 2018; Pastor 2015], укажем отличительные черты словаря Antidote.

Прежде всего следует отметить, что здесь последовательно используются корпусные данные, поэтому есть возможность получать статистическую информацию. Кроме того, благодаря использованию корпусных данных словарь позволяет представлять большое количество сочетаний (для наиболее частотных лексем это количество может достигать нескольких сотен и даже тысяч; так, для французского глагола *aller* приведено 1027, а для английского *to go* – 2345 сочетаний). При помощи реализованной в словаре функции поиска пересечений сочетания могут рассматриваться в сходных контекстах, что позволяет анализировать случаи синонимии на синтаксическом уровне. В целом благодаря тому, что слова сопровождаются контекстами в виде большого количества предложений, в которых они выполняют различные синтаксические функции, данный словарь можно назвать «словарем предложений».

Сильные стороны словаря, превращающие его в мощный инструмент для теоретического и практического исследования языка, однако, в определенной мере становятся его недостатками. Одним из самых существенных является то, что использованные в словаре корпусные данные не вполне сбалансированы. Во французскую часть словаря недифференцированно входят художественные тексты (причем многие из них переводные, размещенные на сайте Проекта Гутенберг), научная литература, публицистика, статьи из Википедии и т. п., – такие тексты соответствуют книжной литературной норме, но в некотором отношении ее сужают. Речь в данном случае идет не о том, что в словаре представлена стилистически однородная лексика, а о том, что в качестве иллюстративного материала используется преимущественно книжный вариант языка с присущими ему грамматическими формами и синтаксическими конструкциями.

По-видимому, методами автоматического вывода сочетаний обусловлен тот факт, что в разделе сочетаний отсутствует разделение между свободными и связанными, фразеологизированными сочетаниями.

Кроме того, следует отметить еще один момент, обусловленный ориентацией на корпусные данные, который может вызвать трудности у пользующихся данным словарем как учебным, – относительную сложность предложений, выводимых в разделе сочетаний: предложения аутентичны и поэтому не адаптированы для изучающих язык.

Перспективы учебной лексикографии в свете корпусной методологии

Проведенный анализ словарей и программных средств, ориентированных на «активное» изучение языка, был выполнен с целью сравнить методические возможности традиционных и относительно новых – электронных – источников информации о языке. В ходе анализа было установлено, что в словарях активного типа, ориентированных на практическое овладение языком, в неявном виде преобладает лексико-ориентированный подход: в центре словарной статьи оказывается либо слово, либо словосочетания с этим словом; предложениям же уделяется внимание лишь постольку, поскольку предложения способны раскрыть значение слова в контексте или проиллюстрировать его употребление. Ни в коем случае не утверждая, что такой подход является устаревшим, считаем важным отметить, что современные средства автоматической обработки естественного языка позволяют существенно расширить потенциал словарей активного типа. Корпусные данные открывают возможность с минимальными затратами получить гораздо более широкую и актуальную картину сочетаемости особенностей запрашиваемых лексем.

Статистическая информация (которая в явном виде, как правило, отсутствует в словарях сочетаемости) при этом играет чрезвычайно важную роль в описании лексического значения слова: наиболее частотная дистрибуция указывает на то, в каких значениях лексема употребляется чаще (об анализе дистрибуции как средстве снятия многозначности см., например: [Atkins, Rundell 2008: 294 et seq.]), а это обязательно учитывается при определении порядка лексических значений в статьях лингвистических словарей, в особенности если эти словари имеют учебную направленность.

К сожалению (и это было показано на отдельных примерах запросов в BNC), корпусные менеджеры позволяют управлять выдачей результатов в большинстве случаев только с помощью ввода чисто морфологических признаков (те возможности – весьма, впрочем, небольшие – для указания семантических признаков в запросе, которые предоставляет НКРЯ, являются скорее исключением, чем правилом). В результате запросов пользователь получает данные, которые невозможно упорядочить семантически, что, по-видимому, обусловлено сложностями, связанными с формализацией семантики естественных языков. Такая же ситуация вывода семантически неупорядоченного списка сочетаний наблюдается в словаре Antidote.

Тем не менее получаемая при формировании запросов картина сочетаемости лексем позволяет с некоторыми интеллектуальными усилиями реконструировать модель их управления, особенности распространения и наиболее важные синтаксические функции.

В связи с моделями управления и синтаксическими функциями необходимо сделать оговорку, что среди приведшихся примеров только данные электронного словаря Antidote позволяли группировать сочетания в соответствии с тем, какими словами управляет и какую функцию в предложении выполняет запрашиваемая лексема. Подобная группировка возможна не во всех корпусных менеджерах, однако в определенной степени она может быть достигнута, если при создании запроса будут корректно указаны морфологические признаки запрашиваемых словоформ и их позиции относительно друг друга.

Тем не менее приходится констатировать, что синтаксический анализ, который позволял бы представлять структуру предложений с запрашиваемыми лексемами в виде дерева зависимостей (или ином синтаксически структурированном виде), в большинстве современных корпусных менеджеров отсутствует. Из находящихся в свободном доступе известных примеров корпусов русского языка можем указать синтаксически размеченный корпус в НКРЯ, насчитывающий 1,4 миллиона слов,

Отсутствие удовлетворительного семантического и синтаксического анализов усложняют взаимодействие с корпусными банками данных – в то время как в классических словарях этот анализ «вручную» проделывают их составители; с другой стороны, объем информации, который может быть извлечен из корпусных данных, в сотни раз превосходит объем, который можно включить в учебный и даже академический лингвистический словарь.

Очевидно, устранению данного несоответствия между возможностями корпусных данных и классических словарей могут способствовать развитие информационных технологий и активное применение их в лингвистике, однако перспектива такого решения проблемы для академической лексикографии представляется пока весьма отдаленной. В учебной же лексикографии, где объемы данных значительно меньше, можно наметить путь, двигаясь по которому можно относительно легко создать современные, соответствующие требованиям цифровой эпохи справочные издания.

На наш взгляд, возможна разработка электронных обучающих словарей на основе корпусных данных, при этом должен использоваться обучающий – не учебный – корпус, который будет представлять язык в «адаптированном», предназначенном для изучения русского языка как иностранного виде (в английской терминологии «*pedagogic corpus*» [Baker 2006: 128]). В качестве источников для подобного корпуса можно использовать адаптированные тексты на русском языке, тексты учебных пособий, словарей и справочников по РКИ. Похожий корпус, включающий тексты учебников по РКИ, разработан сотрудниками Государственного института русского языка им. А. С. Пушкина [Лапошина 2020], однако он нацелен на решение иных задач, а именно – задач, связанных с оценкой представленного в учебниках языкового материала.

На основе сформированного корпуса адаптированных текстов можно разработать систему поиска, подобную корпусному менеджменту, которая будет позволять создавать поисковые запросы синтаксического характера так, чтобы пользователь имел возможность получать информацию о различных синтаксических функциях запрашиваемых лексем, их сочетаемостных особенностях и т. п. – и все это в пределах соответствующего уровня владения языком с примерами предложений, легко поддающихся пониманию и выявлению лежащих в их основе моделей. Кроме того, в созданном таким образом небольшом корпусе возможно будет произвести тегирование синтаксических структур, которое позволит создавать чисто синтаксические, структурные запросы (типа N1 Vf N4, где N1 – существительное / местоимение / субстантиват в И. п. в функции подлежащего, Vf – спрягаемая форма глагольного сказуемого, N4 – прямое дополнение, выраженное существительным в В. п.); при этом должна быть реализована возможность с помощью семантических тегов задавать параметры компонентов синтаксической структуры, указывать тип действия, если предикат обозначает действие, семантический тип актанта, тип сирконстанта, если предложение распространено обстоятельством и т. п.

На наш взгляд, охарактеризованный выше электронный источник информации будет обладать большой ценностью для изучающих РКИ, т. к. позволит сделать шаг от слова и словосочетания к предложению. Любой естественный язык в своем натуральном, «необработанном» состоянии существует не в виде словарей и грамматик, а в виде предложений, поэтому создание электронного банка данных адаптированных текстов на русском языке и системы управления им – своеобразного «словаря предложений» – станет попыткой описания части языка, по возможности максимально репрезентирующей целое, в наиболее естественном его виде.

ИСТОЧНИКИ

Леонтьева, Т. В. Словарь актуальной лексики единения и вражды в русском языке начала XXI века / Т. В. Леонтьева, А. В. Щетинина. – Екатеринбург : Ажур, 2021. – 424 с.

Новый большой англо-русский словарь : в 3 т. / Ю. Д. Апресян [и др.] ; под общ. рук. Ю. Д. Апресяна. – М. : Русский язык, 1993–1994.

- Русский семантический словарь. Толковый словарь, систематизированный по классам слов и значений / под общ. ред. Н. Ю. Шведовой : в 4 т. Т. I. – М. : Азбуковник, 2002. – ххiii, 803 с.
- Словарь русского языка коронавирусной эпохи / ред. Е. С. Громенко, А. С. Павлова, М. Н. Приемышева [и др.]. – СПб. : Институт лингвистических исследований РАН, 2021. – 550 с.
- Словарь сочетаемости слов русского языка / под ред. П. Н. Денисова, В. В. Морковкина. – М. : Русский язык, 1983. – 688 с.
- Тарасов, Е. Ф. Учебный ассоциативный словарь русского языка / Е. Ф. Тарасов, В. В. Дронов, Е. С. Ощепкова. – СПб. : Златоуст, 2017. – 356 с.
- Толково-комбинаторный словарь русского языка: опыты семантико-синтаксического описания русской лексики / сост. и ред. А. Жолковский, И. Мельчук. – 2-е изд. испр. и доп. – М. : Языки славянской культуры, 2016. – 544 с.
- Benson, M. The BBI Combinatory Dictionary of English / M. Benson, E. Benson, R. Ison. – Amsterdam ; Philadelphia : John Benjamins Publishing Company, 2010. – xli, 462 p.
- Dictionnaire des combinaisons de mots / sous la direction de D. Le Fur. – Paris : Le Robert, 2007. – viii, 1011 p.
- Longman Collocations Dictionary and Thesaurus / K. Cleveland-Marwick, L. Hollingworth, E. Manning et al. – Harlow : Pearson Education Limited, 2013. – vii, 1463 p.
- Macmillan Collocations Dictionary for Learners of English / M. Rundell. – Macmillan Publishers Limited, 2010. – xiii, 911 p.
- Oxford Collocations Dictionary: for Students of English / M. Runcie. – Oxford : Oxford University Press, 2002. – xiii, 897 p.
- Sharoff, S. Frequency Dictionary of Russian: Core Vocabulary for Learners / S. Sharoff, E. Umanskaya, J. Wilson. – London : Routledge, 2013. – viii, 384 p.

ЛИТЕРАТУРА

- Влавацкая, М. В. Комбинаторная лингвистика: аспекты изучения сочетаемости слов : монография / М. В. Влавацкая. – Новосибирск : НГТУ, 2016. – 244 с.
- Воронцова, И. А. Современные практики англоязычной онлайн-лексикографии / И. А. Воронцова // Верхневолжский филологический вестник. – 2022. – № 1 (28). – С. 186–194.
- Захаров, В. П. Корпусная лингвистика : учебник / В. П. Захаров, С. Ю. Богданова. – 3-е изд., перераб. – СПб. : Изд-во С.-Петерб. ун-та, 2020. – 234 с.
- Лапошина, А. Н. Корпус текстов учебников РКИ как инструмент анализа учебных материалов / А. Н. Лапошина // Русский язык за рубежом. – 2020. – № 6 (283). – С. 22–28.
- Лебедева, М. Ю. Дано мне тело – что мне делать с ним? Применение корпусных технологий в лингводидактике РКИ / М. Ю. Лебедева // Русский язык за рубежом. – 2020. – № 6. – С. 4–13.
- Перспектив активного словаря русского языка / В. Ю. Апресян, Ю. Д. Апресян, Е. Э. Бабаева [и др.] ; отв. ред. акад. Ю. Д. Апресян. – М. : Языки славянских культур, 2010. – 784 с.
- Русский язык коронавирусной эпохи : коллективная монография / Е. С. Громенко, Н. В. Козловская, А. С. Павлова [и др.]. – СПб. : Институт лингвистических исследований РАН, 2021. – 610 с.
- Хохлова, М. В. Исследование лексико-семантической сочетаемости в русском языке с помощью статистических методов : автореф. дис. ... канд. филол. наук / Хохлова М. В. – СПб., 2010. – 26 с.
- Atkins, B. T. Sue. The Oxford Guide to Practical Lexicography / B. T. Sue Atkins, M. Rundell. – Oxford : Oxford University Press, 2008. – xii, 540 p.
- Baker, P. Glossary of Corpus Linguistics / P. Baker, A. Hardie, T. McEnery. – Edinburgh : Edinburgh University Press, 2006. – 187 p.
- Dziemiainko, A. Electronic Dictionaries / A. Dziemiainko // The Routledge Handbook of Lexicography / edited by Pedro A. Fuertes-Olivera. – London ; New York : Routledge, 2018. – P. 663–682.
- Guide d'utilisation d'Antidote 11. – URL: <https://www.antidote.info/user/data/guides/antidote-11-documentation-guide-utilisation-fr.pdf> (mode of access: 25.08.2022). – Text : electronic.
- Kilgarriff, A. Introduction to the Special Issue on the Web as Corpus / A. Kilgarriff, G. Grefenstette // Computational Linguistics. – 2003. – No. 29 (3). – P. 333–347.
- Kilgarriff, A. Using Corpora as Data Sources for Dictionaries / A. Kilgarriff // The Bloomsbury Handbook of Lexicography / edited by H. Jackson. – London : Bloomsbury, 2022. – P. 71–88.
- Kormacheva, D. Evaluation of Collocation Extraction Methods for the Russian Language / D. Kormacheva, L. Pivovarova, M. Kopotev // Quantitative Approaches to the Russian Language. – Abingdon : Routledge, 2018. – P. 137–157.
- Kosem, I. Interrogating a Corpus / I. Kosem // The Oxford Handbook of Lexicography / edited by P. Durkin. – Oxford : Oxford University Press, 2016. – P. 76–93.
- Meurers, D. W. Corpora and Syntax / D. W. Meurers, S. Müller // Corpus Linguistics: An International Handbook. Vol. 2 / edited by A. Lüdeling, M. Kytö. – Berlin ; New York : Walter de Gruyter, 2009. – P. 920–933.
- Pastor, V. Researching the Use of Electronic Dictionaries / V. Pastor, A. Alcina // The Bloomsbury Handbook of Lexicography / edited by H. Jackson. – London : Bloomsbury, 2022. – P. 89–130.
- Stefanowitsch, A. Corpus Linguistics: A Guide to the Methodology / A. Stefanowitsch. – Berlin : Language Science Press, 2020. – x, 494 p.

REFERENCES

- Apresyan, V. Yu., Apresyan, Yu. D., Babaeva E. E. et al. (2010). *Prospekt aktivnogo slovarya russkogo yazyka* [Prospectus of an Active Dictionary of Russian]. Moscow, Yazyki slavyanskikh kul'tur. 784 p.
- Atkins, B. T. Sue, Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford, Oxford University Press. 540 p.

- Baker P., Hardie A., McEnery T. (2006). *Glossary of Corpus Linguistics*. Edinburgh, Edinburgh University Press. 187 p.
- Dziemiánko, A. (2018). Electronic Dictionaries. In Fuentres-Olivera, Pedro A. (Ed.). *The Routledge Handbook of Lexicography*. London, New York, Routledge, pp. 663–682.
- Gromenko, E. S., Kozlovskaya, N. V., Pavlova, A. S. et al. (2021). *Russkii yazyk koronavirusnoi epokhi* [The Russian Language of the Coronavirus Era]. Saint Petersburg, Institut lingvistsicheskikh issledovaniy RAN. 610 p.
- Guide d'utilisation d'Antidote* 11. URL: <https://www.antidote.info/user/data/guides/antidote-11-documentation-guide-utilisation-fr.pdf> (mode of access: 25.08.2022).
- Khokhlova, M. V. (2010). *Issledovanie leksiko-semanticheskoi sochetaemosti v russkom yazyke s pomoshch'yu statisticheskikh metodov* [Research of Lexical-Syntactic Collocation in Russian Using Statistical Methods: Corpus-based Study]. Avtoref. dis. ... kand. philol. nauk. Saint Petersburg. 26 p.
- Kilgarriff, A. (2022). Using Corpora as Data Sources for Dictionaries. In Jackson, H. (Ed.). *The Bloomsbury Handbook of Lexicography*. London, Bloomsbury, pp. 71–88.
- Kilgarriff, A., Grefenstette, G. (2003). Introduction to the Special Issue on the Web as Corpus. In *Computational Linguistics*. No. 29 (3), pp. 333–347.
- Kormacheva, D., Pivovarova, L., Kopotev, M. (2018). Evaluation of Collocation Extraction Methods for the Russian Language. In *Quantitative approaches to the Russian language*. Abingdon, Routledge, pp. 137–157.
- Kosem, I. (2016). Interrogating a Corpus. In Durkin, P. (Ed.). *The Oxford Handbook of Lexicography*. Oxford, Oxford University Press, pp. 76–93.
- Laposhina, A. N. (2020). Korpus tekstov uchebnikov RKI kak instrument analiza uchebnykh materialov [A Corpus of Russian Textbook Materials for Foreign Students as an Instrument of an Educational Content Analysis]. In *Russkii yazyk za rubezhom*. No. 6, pp. 22–28.
- Lebedeva, M. Yu. (2020). Dano mne telo – chto mne delat' s nim? Primenenie korpusnykh tekhnologii v lingvodidaktike RKI [I'm Given a Corpus – What to Do With It? Corpus Technologies in Russian Language Teaching and Learning]. In *Russkii yazyk za rubezhom*. No. 6, pp. 4–13.
- Meurers, D. W., Müller, S. (2009). Corpora and Syntax. In Lüdeling, A., Kytö, M. (Eds.). *Corpus Linguistics: An International Handbook*. Vol. 2. Berlin, New York, Walter de Gruyter, pp. 920–933.
- Pastor, V., Alcina, A. (2022). Researching the Use of Electronic Dictionaries. In Jackson, H. (Ed.). *The Bloomsbury Handbook of Lexicography*. London, Bloomsbury, pp. 89–130.
- Stefanowitsch, A. (2020). *Corpus Linguistics: A Guide to the Methodology*. Berlin, Language Science Press. 494 p.
- Vlavatskaya, M. V. (2016). *Kombinatornaya lingvistika: aspekty izucheniya sochetaemosti slov* [Combinatorial Linguistics. Aspects of Word Combinatorial Power Study]. Novosibirsk, NGTU. 244 p.
- Vorontsova, I. A. (2022). Sovremennye praktiki angloyazychnoi onlain-leksikografii [Modern Practices of English Online Lexicography]. In *Verkhnevolzhskii filologicheskii vestnik*. No. 1, pp. 186–194.
- Zakharov, V. P., Bogdanova, S. Yu. (2020). *Korpusnaya lingvistika* [Methods in Corpus Linguistics]. Saint Petersburg, Izdatel'stvo Sankt-Peterburgskogo un-ta. 234 p.

Данные об авторе

Гончаренко Илья Георгиевич – кандидат филологических наук, доцент,
Российский государственный профессионально-педагогический университет
(Екатеринбург, Россия).

Адрес: 620143, Россия, Екатеринбург, ул. Машиностроителей, 11.

E-mail: ilya.goncharenko@rsvpu.ru.

Author's information

Goncharenko Ilia Georgievich – Candidate of Philology, Associate
Professor, Russian State Vocational Pedagogical University (Ekaterinburg,
Russia).

Дата поступления: 28.08.2022; дата публикации: 30.06.2023

Date of receipt: 28.08.2022; date of publication: 30.06.2023