

Сайда Биктимировна ХАБИБУЛЛИНА¹
Ольга Борисовна УЛЬЯНОВА²

УДК 82-193

КОРПУСНЫЙ АНАЛИЗ ПРЕДИКАТОВ ПЕРЕДАЧИ ЗНАНИЯ В ТЕКСТАХ АННОТАЦИЙ К НАУЧНЫМ СТАТЬЯМ

¹ старший преподаватель
кафедры английской филологии и перевода,
Тюменский государственный университет
s.b.khabibullina@utmn.ru

² кандидат филологических наук, доцент
кафедры английского языка,
Тюменский государственный университет
o.b.ulyanova@utmn.ru

Аннотация

В статье методы корпусной лингвистики применяются для исследования семантики общеначальных глаголов лексико-семантической группы *передачи знания* с целью изучения семантической организации и тематического упорядочивания данной группы англоязычных предикатов в тексте аннотации. Категориальное таксономическое значение *передачи знания* обеспечивает адекватное восприятие информации при сжатии основного текста научной статьи. Существующие исследования данной направленности посвящены риторической структуре, или лингвокогнитивной организации аннотаций к научным статьям в различных предметных областях. В рамках аннотаций парадигматика и синтагматика лексических единиц в целом и предикатов в частности остаются не в полной мере изученными. Следовательно, актуальность предмета исследования, а именно предикатов *передачи знания* в текстах аннотаций, обусловлена объективной необходимостью выполнения коммуникативной задачи создания или перевода текста аннотации преимущественно с русского на английский. При этом неанглоязычное

Цитирование: Хабибуллина С. Б. Корпусный анализ предикатов *передачи знания* в текстах аннотаций к научным статьям / С. Б. Хабибуллина, О. Б. Ульянова // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. 2020. Том 6. № 1 (21). С. 62-75.

DOI: 10.21684/2411-197X-2020-6-1-62-75

академическое сообщество нуждается в доступе к аутентичным изысканиям, понимание которых происходит преимущественно на основании предикатов пропозиции. На материале корпуса, созданного из 500 текстов аннотаций к статьям предметной области лингвистика посредством применения автоматизированного количественного и качественного методов корпусного анализа, производится отбор предикатов и формируется лексико-семантическая группа *передачи знания* с семантической доминантой *to show*, которая проявляет самую высокую частотность употребления в текстах англоязычной аннотации. Наряду с ядерной семой *показа* выделяются семы: *появление знания*, *подтверждение знания*, *разъяснение знания*, *акцентуация знания*, *обзор знания*, которые организуют пространство лексико-семантической группы *передачи знания* и, следовательно, тексты аннотаций. Синтагматика исследуемых глаголов сводится к четырем типам моделей сочетания глагола и прямого дополнения; глагола и предложного дополнения; глагола и придаточного предложения, а также глагола и инфинитива, где первая модель наиболее частотна, а последняя наименее частотна.

Ключевые слова

Корпусный анализ, лексико-семантическая группа, сочетаемость, частотность словоупотреблений.

DOI: 10.21684/2411-197X-2020-6-1-62-75

Введение. Постановка проблемы

Формирование единого академического пространства накладывает определенные требования на организацию текста научной статьи. Установки на доступность и открытость научных исследований ориентируют ученых делиться изысканиями преимущественно на английском языке. В эпоху информационной этики, безопасности и лицензирования авторского права доступ к полнотекстовым научным статьям, документам и материалам в глобальной сети Интернет не всегда возможен. В открытом доступе в ряде научных журналов оказываются только тексты аннотаций, которые являются кратким представлением того, что описано в первоисточнике. Именно поэтому изучение текстов аннотаций является востребованным, аннотация помогает определить релевантность той или иной работы и повышает интерес к научной статье. Актуальность анализа состоит в том, что даже в российском научном сообществе статье предшествует не только русскоязычная, но и англоязычная аннотация.

Настоящее исследование глаголов *передачи знания* выполнено в рамках корпусной лингвистики и автоматизированного количественного отбора и анализа для дальнейшей качественной классификации и использования данного корпуса в практике обучения, создания текстов аннотаций или их перевода (преимущественно с русского на английский) в рамках неанглоязычного студенческого и научного сообществ. В качестве объекта изучения выступает созданный для проведения исследования корпус текстов аннотаций, а предмета — общенаучные предикаты, используемые в данном типе текста. Цель ис-

следования заключается в семантической организации и тематическом упорядочивании английских глаголов с общей семой *передача знания* в тексте аннотации к научной статье.

Современная лингвистика располагает комплексным инструментарием для оптимизации и совершенствования процесса лингвистического анализа и увеличения объема обрабатываемых данных [3, с. 47-50] с помощью компьютеров и электронных корпусов [4, с. 48-59]. Будучи направленным на использование средств языка как инструмента создания текстов аннотаций (обработку данных естественного языка), исследование опирается на понятие корпуса [21, 17, 22] для создания собственного корпуса англоязычных аннотаций к научным статьям и использования автоматических функций корпус-менеджера для количественных (статистических или выявления частотности) и качественных (классификационных и интерпретативных) методов [7, с. 104-107] морфологического и синтаксического анализа текстов [2, с. 135-137], их разметки при помощи лексико-грамматических тегов, выявления случаев сочетаемости ключевых единиц посредством конкорданса с целью отбора и полиаспектного описания глаголов лексико-семантической группы *передачи знания*.

Целостная семантическая структура лексической единицы (в настоящей статье — глагола) складывается из архисемы и дифференциальных сем, различающихся по ряду признаков, что позволяет выделить некоторые тематические группы и тем самым упорядочить единицы языка. Семантическая природа глагола соотносится с двумя компонентами: грамматическим и лексическим [6, с. 502]. Грамматической особенностью глагольной семантики является отражение в значении глагола признаков состояния, процессуальности, посессивности, наличие грамматических форм и широкий сочетаемостный потенциал [6, с. 504]. В качестве одного из целесообразных способов фиксирования семантического соседства глаголов предлагается изучение лексико-семантического статуса рассматриваемых единиц внутри организованной группы (понятийной, тематической или лексико-семантической). Лексемы способны структурировать контекстуальное пространство, а корпусные методы обработки текста призваны автоматически установить регулярную сочетаемость анализируемого слова в его определенном лексическом значении в отрыве от полисемии и соотнести элементы с тематическими классами [9, с. 216].

Начиная с периода расцвета корпусной лингвистики в 1980-х гг. [10] и по настоящее время предпринят ряд исследований, к примеру, текстов аннотаций с точки зрения их риторической организации [13, 20], макро-, микроструктуры и метадискурса [19], сопоставительного анализа стратегий в английском и китайском языках [12], корпуса предикатов [15, 16], классификации предикатов [18, 8], семантики предикатов [1] и пропозиций [14].

Следуя методике представления лингвистического корпуса [2, с. 135-137], настоящий корпус аннотаций реализован в машиночитаемом формате, структурирован и размечен для решения лингвистической задачи [5, с. 5] выявления синтаксики и семантики предикатов *передачи знания*. Мы объединяем все ме-

тадискурсивные глаголы, которые несут в себе новое или существующее знание, в предикаты передачи знания, поскольку любое исследование, отраженное в научной статье, связано с представлением, поиском и обзором информации или знания. К лексико-семантической группе глаголов *передачи знания* мы относим глаголы, которые выполняют метаязыковую функцию в научном тексте.

Очевидно, что текст аннотации построен следуя всем принципам организации текста и логично, целостно и лаконично отражает научные достижения [23], представленные в виде статьи, тезисов конференции или иных научных материалов, выступая сокращенной версией развернутого формата научной статьи, позволяя заинтересованной аудитории исследователей избирательно подойти к отбору нужных изысканий.

Глаголы *передачи знания* (как основы пропозиции научного текста) способствуют концентрации на определенной тематике, помогают эксплицировать суть высказывания и интерпретировать необходимые данные. Следовательно, полагаем, что семантика предикатов должна соответствовать принципам генерализации, специализации, презентации, экземплификации, экспликации и интерпретации содержания статьи. Таким образом, мы предпринимаем попытку описания не только статистических результатов, но и синергии естественного языка, научного дискурса и семантики.

Проведение корпусного исследования

Исследование текстов аннотаций к статьям по лингвистике проводилось в определенной последовательности действий. Для начала был сформирован подкорпус из 500 аннотаций на английском языке, размещенных в англоязычных научных журналах по лингвистике (*Journal of Linguistics*, *Applied Linguistics* и *Journal of English Linguistics* с 2008 по 2015 г.). Параметрами отбора аннотаций, призванными обеспечить достоверность данных и большую объективность проведения исследования, послужили: импакт-фактор журнала (по данным сайта scimago.com), предметное лингвистическое соответствие журнала, предметное лингвистическое наполнение статьи, дата публикации, редакторы — носители английского языка.

Необходимость создания собственного корпуса была обусловлена тем, что отдельно тексты аннотаций не входят в известные корпусы. Так, пятьсот аннотаций были собраны в корпус посредством приложения — текстового редактора «Блокнот» — и сохранены при помощи текстового формата *txt и кодировки UTF-8. Объем исследуемого файла составляет 1,16 МБ (1 226 455 байт). Из множества лингвистических компьютерных программ, способствующих описанию корпуса текстов, данный корпус описан при помощи программы *AntConc**, технические характеристики которой удовлетворяют требованиям работы в системе *Windows*.

Второй этап настоящего анализа заключается в разметке неразмеченного корпуса аннотаций с помощью программы *TreeTagger* (автоматизированного инструмента морфологической разметки текста, включающего 58 тегов), а

именно функции морфологического дескриптора для обозначения частей речи, для отбора глаголов. Рис. 1 содержит размеченный корпус текстов аннотаций.

Из набора тегов TreeTagger Tag Set 6 тегов частеречной разметки актуальны для данного исследования: VVZ (verb, present 3^d person singular), VVP (verb, present, non-3^d person), VVD (verb, past tense), VVN (verb, past participle), VVG (verb, gerund/participle). Использование разметки для корпуса из 500 аннотаций позволило выявить 3 353 предиката из общего количества 203 915 лексических единиц.

На третьем этапе исследования посредством семантического анализа была сформирована лексико-семантическая группа с семантической доминантой «передача знания». Ее объем составил 60 единиц. В лексико-семантической группировке константной является семантика *передачи знания*, а вариативность представлена семами *демонстрации знания, выявления, обоснования, разъяснения, уточнения и систематизации знаний*. Автоматическое составление конкорданса дает возможность систематизации и активации тематически упорядоченной лексики. Список наиболее употребительных английских общенакальных глаголов с предельно сжатыми контекстами употребления позволяет вывести лексико-грамматические конструкции сочетаемости — абстрактные шаблоны, предполагающие лексикализацию в процессе порождения научного текста. В результате обработки научных аннотаций при помощи инструментария количественного метода был получен список глаголов лексико-семантической группы *передача знания* и выявлено общее количество словоупотреблений (Concordance Hits) данных глаголов в исследуемом типе текста. Применение функции конкорданса для формирования упорядоченного списка словоформ позволило ранжировать глаголы по критерию частотности и упорядочить их по убыванию (далее в списке глаголов в скобках дается частотность их употребления), указав на первом месте самый популярный глагол лексико-семантической группы *передача знания*: *to show* (330), *to argue* (224), *to present* (90), *to examine* (88), *to follow* (72), *to appear* (58), *to explain* (54), *to explore* (53), *to indicate* (50),

1	NN support and/or CC and/or accentuate	VV accentuate contrast NN contrast between IN before	ADSL
2	' JJ ' ability NN ability to TO to access	W access multiple JJ multiple interpretations NNS i	Abstr
3	then modified VVN modify to TO to accommodate	W accmmodate the DT the nonverbal JJ nonverbal app	
4	identity choices NNS choice to TO to accommodate	W accmmodate to TO to one CD one group	Abstr
5	ferent resources NNS resource to TO to accomplish	W accomplish this DT this . SENT . Moreover RB more	
6	be extended VVN extend to TO to account	W account for IN for certain JJ certain asymmetrica	Abstr
7	mechanism fail WP fail to TO to account	W account for IN for synchronic JJ synchronic and	Abstr
8	alone does VZ do not RB not account	W account for IN for all DT all postnasal	Abstracts_2008_2015ta
9	allow us PP us to TO to account	W account for IN for a DT a number	Abstracts_2008_2015tag.txt
10	then used VVN use to TO to account	W account for IN for many DT many hitherto	Abstracts_2008_2015ta
11	terpretations NNS interpretation to TO to account	W account for IN for the DT the semantic	Abstr
12	that one PP one cannot MD can account	W account for IN for the DT the felicity	Abstracts_200
13	be able JJ able to TO to account	W account for IN for the DT the full	Abstracts_2008_2015tag.txt

Рис. 1. Пример размеченного корпуса текста аннотации на английском языке

Fig. 1. Part of an Abstract with POS Tagging

to demonstrate (48), to draw (42), to reveal (38), to determine (35), to exhibit (32), to emerge (31), to express (31), to mark (28), to represent (28), to observe (27), to report (26), to read (25), to establish (23), to write (23), to illustrate (21), to produce (21), to reflect (21), to create (20), to outline (17), to highlight (16), to review (16), to confirm (13), to map (13), to prove (13), to perform (12), to display (10), to publish (10), to attest (9), to point (9), to promote (7), to assert (6), to discover (4), to project (4), to render (4), to clarify (3), to evidence (3), to expose (3), to note (3), to record (3), to uncover (3), to chart (2), to expand (2), to feature (2), to instruct (2), to substantiate (2), to depict (1), to manifest (1), to obey (1), to print (1), to unfold (1), to witness (1).

Список демонстрирует, что самыми употребительными глаголами являются *to show* (330) и *to argue* (224), чуть реже используются глаголы *to present* (90), *to examine* (88) и *to follow* (72). Кроме того, на наш взгляд, нужно отметить и остальные глаголы списка, которые употребляются в корпусе не менее чем 15 раз (первые 30 глаголов). Они могут служить синонимами для вышеуказанных самых частотных глаголов, позволяя авторам полнее и точнее выражать их идеи в тексте аннотации к научной статье. По мере снижения частотности разрыв в количестве употреблений снижается и составляет не более трех пунктов.

На предпоследнем этапе данного исследования, в целях тематического упорядочивания английских глаголов исследуемой лексико-семантической группы, полученных в результате корпусного исследования, нам удалось вычленить некоторые семантические категории. Как было описано выше, посредством семантического анализа были отобраны общенаучные глаголы с общей семантикой *передачи знания*, которые принадлежат одной лексико-семантической группе, поскольку в семантической структуре данных глаголов, по данным функции тезауруса словаря Cambridge Dictionary Online [11], на первый план выходит категориальная сема / архисема «передача знаний» (в тексте научной аннотации). Однако значения всех исследуемых глаголов можно конкретизировать наличием как интегральных, так и дифференциальных сем.

Вслед за попытками ученых классифицировать общенаучные глаголы *передачи знания* (к примеру, [1]) мы выявили шесть понятийных групп глаголов внутри лексико-семантической группы *передачи знания*. Для текстов аннотаций, по материалам корпуса, характерно использование следующей понятийной группы, состоящей из 26 глаголов: *to show, to present, to indicate, to demonstrate, to draw, to illustrate, to exhibit, to attest, to express, to display, to mark, to represent, to establish, to reflect, to outline, to map, to point, to chart, to depict, to write, to publish, to read, to record, to print, to instruct, to observe*. Семантической доминантой, или ядром данной группы глаголов является глагол *to show*. Семантика «показа» выступает связующим звеном в тексте аннотации, тогда как отдельные предикаты могут отличаться друг от друга оттенками значения. К примеру, для визуализации и экземплификации проблематики используются предикаты *to depict, to illustrate, to outline*, а для представления ее в определенном графическом формате употребляются глаголы *to chart, to map*. В свою

очередь, глагол *to display* обозначает показ с оттенком обнаружения, глаголы *to present, to represent* представляют объект на обозрение, глагольные единицы *to exhibit / to demonstrate* делают акцент на развертывании проблемы или понятия, предикаты *to point / to instruct* имеют значение показа через четкое указание. Для того, чтобы уточнить письменную/печатную форму демонстрации исследуемого материала, в ряде случаев используются глаголы *to write, to publish, to print*.

Вторая понятийная группа насчитывает 14 глаголов: *to reveal, to appear, to emerge, to create, to produce, to perform, to project, to expose, to uncover, to manifest, to unfold, to follow, to obey, to discover*, которые объединены, как и все глаголы анализируемой лексико-семантической группы, семой передачи знания, но, в отличие от первой понятийной группы, связующим элементом выступает не *показ* в разных формах исполнения, а *обнаружение и появление/возникновение* нового знания, мыслей и идей.

Английские глаголы *to confirm, to prove, to promote, to assert, to evidence, to substantiate, to witness* (в рамках третьей по численному составу группы — 7 предикатов) несут в себе семантику передачи знания через обоснование, доказательство и подтверждение какого-то знания.

Четвертая группа включает 6 глаголов: *to explain, to argue, to clarify, to expand, to determine, to report*, которые, с одной стороны, сообщают информацию, а с другой стороны, уточняют, разъясняют, определяют или расширяют условия и обстоятельства, при которых проходит исследование или наблюдение.

Пятая группа объединяет глаголы *to highlight, to render, to note, to feature*, которые передают знание через акцентуацию, то есть выделение, отдельных элементов знания, как правило, особо значимых для автора в проводимом исследовании.

Шестая понятийная группа глаголов *передачи знания* указывает на тщательное обзорное изучение при помощи таких глагольных единиц, как *to examine, to explore, to review*.

Итак, представленные понятийные группы реализуют семантику *сообщения* или *передачи знаний*, уточняя их отдельными компонентами значения. Так, нам удалось выделить следующие концептуальные области: *показ, демонстрация знания; обнаружение, появление знания; обоснование, подтверждение знания; разъяснение знания; акцентуация знания; обзор знания*.

Диаграмма на рис. 2 представляет обзор концептуальной организации текста аннотации посредством глаголов, активирующих различные зоны текста аннотации.

Последний этап исследования, проведенный посредством качественного метода корпусного анализа, демонстрирует в узком смысле лексико-грамматическую сочетаемость анализируемых глаголов через правые коллокации данных глаголов, а в широком смысле вскрывает глубинную семантическую структуру пропозиции *передача знания*. С помощью функции программы AntConc «Advanced Search» все глаголы были внесены в список корпуса и проведен количественный анализ правых коллокаций исследуемых глаголов (рис. 3).



Рис. 2. Концептуальные области применения исследуемых глаголов

Fig. 2. Conceptual Areas of Applying Reporting Verbs

Результаты данного этапа корпусного исследования помогли установить через правые коллокации данных глаголов, что глаголы описываемой лексико-семантической группы имеют общие синтагматические характеристики, употребляются в одинаковых типах синтаксических структур, имеют общие семантические роли аргументов и сочетаются с одними и теми же разрядами существительных. Базовые лексико-грамматические конструкции, по которым можно установить лексико-грамматическую сочетаемость данных глаголов: «глагол + существительное (прямое дополнение)»; «глагол + существительное (предложное дополнение)»; «глагол + придаточное предложение»; «глагол + инфинитив».

Интерпретация правого развертывания коллокаций указывает на то, что самой частотной конструкцией в аспекте лексико-грамматической сочетаемости глаголов *передачи знания* стала модель «глагол + существительное в функции прямого

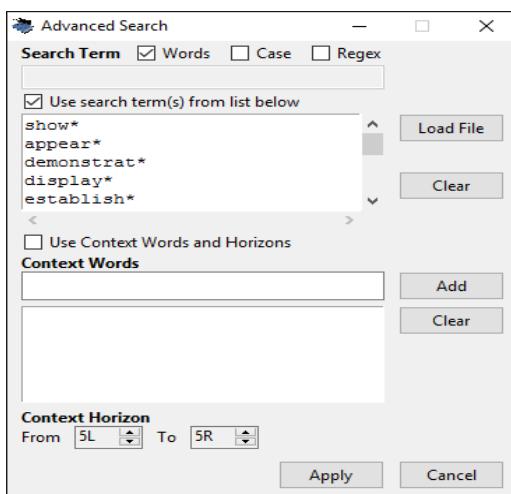


Рис. 3. Функция «Advanced Search» в программе AntConc

Fig. 3. AntConc Advanced Search Function

дополнения», то есть глагол с объектной семантикой (частотность употребления — 1 080 случаев): *to show systematic phonetic differences; to indicate some theoretical consequences; to chart the diachronic development; to show a high level; to indicate a change of state; to examine a metaphoric chain shift*. На втором месте (с частотой употребления 359 случаев) находится конструкция «глагол + существительное (в функции предложного дополнения)»: *to argue for an alternative; to be marked with respect to; to expand upon analyses; to be shown for language*. Следующей по частотности употребления является модель «глагол + придаточное предложение» (235 случаев): *to illustrate how multiple nominals share a single grammatical function; to demonstrate that each object can bear any of the non-agentive roles; to highlight how experimental methods can inform semantic theory; to show that specific noun phrases are attested as direct objects; to argue that diacritic use of phonological features*. Наименее часто представлена конструкция «глагол + инфинитив» (62 случая употребления): *to be shown to be solely dependent on properties; to appear to be a special case; to be argued to be under the sway of frequency*.

Данное исследование описывает структуру пропозиций с предикатами *передачи знания* в текстах аннотаций к научным статьям. Так, правое развертывание коллокаций позволяет судить о правой части аргумента предиката *передачи знания*. Передача знания по различным каналам и различными способами идентифицируется как направленное действие, объектом которого может выступать любое имя существительное без ограничения (одушевленное/неодушевленное, исчисляемое/неисчисляемое, конкретное/абстрактное), так как любые предметы и явления могут становиться объектами представления и передачи (в рамках лексико-семантической группы *передачи знания*), например, *to show a high level; to demonstrate the crucial role; to map dynamics; to display a systematic alternation; to indicate a change of state; to examine a metaphoric chain shift*. В модели «глагол + существительное (прямое дополнение)» аргументом выступает номинативный компонент с преобладанием количественной семантики степени (a level, a degree) и качественной семантики изменений (a change, a shift, a role, dynamics, differences). Модели «глагол + придаточное предложение» и «глагол + инфинитив» реализуют предикативную сущность правой части аргумента, при этом придаточное предложение преимущественно актуализует глагольный предикат, а инфинитивная конструкция — именной.

Результаты исследования

Корпусный подход к выявлению и изучению языковых особенностей лингвистического содержания текста аннотации на английском языке дал возможность не только систематизировать языковые единицы, но и представить структуру пропозиций элементов анализируемой лексико-семантической группы *передача знания*. Идентификаторами лексико-семантической группы *передача знания* на английском языке стали глаголы *to show, to reveal, to confirm, to explain, to highlight, to examine*, поскольку, с одной стороны, они обладают основными семами, которые имеют глаголы данной группы, а с другой, корпусный анализ

позволил нам установить, что это самые употребительные среди всех общеначальных глаголов (количественный метод). Остальные глаголы исследуемой ЛСГ в своей семантической структуре также имеют соответствующие оттенки значения, но употребляются чуть реже.

Качественный метод корпусного анализа выявил, что глаголы лексико-семантической группы *передача знания* употребляются в одинаковых типах синтаксических структур и что объектом к исследуемым глаголам может выступать любое имя существительное без ограничения.

Так как данные глаголы образуют синонимический ряд в структуре лексико-семантической группы *передача знания*, стоит отметить, что исследуемая лексико-семантическая группа является открытым и незаконченным образованием, способным на дополнения и уменьшения в связи с изменениями, происходящими в контексте или языке.

Подобный анализ общеначальных глаголов призван настроить пользователя, даже не обладающего глубокими знаниями английского языка, использовать самые частотные глаголы с заданной семантикой в своей аннотации, чтобы точно передать свои знания читателю и сделать свой текст максимально доступным для понимания.

СПИСОК ЛИТЕРАТУРЫ

1. Ахманова Г. И. Семантический анализ глаголов со значением передачи знаний в составе научного текста / Г. И. Ахманова, А. М. Вильштейн, М. М. Глушко. М.: МГУ, 1989. 195 с.
2. Баранов А. Н. Введение в прикладную лингвистику: учеб. пособие / А. Н. Баранов. М.: Эдиториал УРСС, 2001. 360 с.
3. Богданова С. Ю. Возможности корпусной методологии в решении лингвистических задач / С. Ю. Богданова // Вестник ИГЛУ. 2012. № 2ю (18). С. 47-50.
4. Захаров В. П. Поисковые системы Интернета как инструмент лингвистических исследований / В. П. Захаров // Русский язык в Интернете: сб. статей. Казань: Отечество, 2003. С. 48-59.
5. Захаров В. П. Корпусная лингвистика / В. П. Захаров, С. Ю. Богданова. СПб.: СПбГУ. РИО. Филологический факультет, 2013. 148 с.
6. Ибрагимова В. Л. Глагол в семантическом пространстве лексики / В. Л. Ибрагимова // Вестник Башкирского университета. 2014. Том 19. № 2. С. 500-505.
7. Лаврентьев А. М. Корпусная лингвистика: идеология, методы, технологии / А. М. Лаврентьев // Сибирский филологический журнал. 2004. № 3-4. С. 121-134.
8. Падучева Е. В. Лексическая аспектуальность и классификация предикатов по Маслову — Вендлеру / Е. В. Падучева // Вопросы языкоznания. 2009. № 6. С. 3-20.
9. Рябцева Н. К. Грамматика конструкций, словари сочетаемости, современные информационные технологии и преподавание иностранных языков / Н. К. Рябцева // Лингвистика и методика преподавания иностранных языков: периодич. сб. науч. ст.,

-
- электронное научное издание. М.: Институт языкоznания РАН, 2014. № 6. С. 214-247. URL: http://www.iling-ran.ru/library/sborniki/for_lang/2014_06/9.pdf (дата обращения: 13.02.2020).
10. Biber D. *Corpus Linguistics: Investigating Language Structure and Use* / D. Biber, S. Conrad, R. Reppen. Cambridge: Cambridge University Press, 1998. 310 p.
 11. Cambridge Dictionary Online. URL: <https://dictionary.cambridge.org/>
 12. Hu G. Hedging and boosting in abstracts of Applied Linguistics articles: a comparative study of English- and Chinese-medium journals / G. Hu, F. Cao // *Journal of Pragmatics*. 2011. Vol. 43. No. 11. Pp. 2795-2809. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0378216611001135>
 13. Hyland K. *Disciplinary Discourses: Social Interactions in Academic Writing* / K. Hyland. London, UK: Longman, 2000.
 14. Kingsbury P. From TreeBank to PropBank / P. Kingsbury, M. Palmer. University of Pennsylvania, 2002. URL: <http://www.lrec-conf.org/proceedings/lrec2002/pdf/283.pdf>
 15. Kipper Schuler K. VerbNet: a Broad-coverage, Comprehensive Verb Lexicon: a dissertation in Computer and Information Science / K. Kipper Schuler. University of Pennsylvania, 2005. URL: <http://verbs.colorado.edu/~kipper/Papers/dissertation.pdf>
 16. Kipper K. A large-scale classification of English verbs / K. Kipper, A. Kornonen, N. Ryant, M. Palmer // *Language Resources and Evaluation Journal*. 2008. No. 42. Pp. 21-40.
 17. Leech G. Corpora and theories of linguistic performance / G. Leech // *Directions in Corpus Linguistics: proceedings of the Nobel Symposium 82* / J. Svartvik (ed.). Berlin; New York: Mouton de Gruyter, 1992. *Trends in Linguistics*. Vol. 65. Pp. 105-122.
 18. Levin B. *English Verb Classes and Alternations: a Preliminary Investigation* / B. Levin. Chicago: The University of Chicago Press, 1993. 366 p.
 19. Ngai S. B. C. A discourse analysis of the macro-structure, metadiscoursal and microdiscoursal features in the abstracts of research articles across multiple science disciplines / S. B. C. Ngai, R. G. Singh, A. C. Koon // *PLoS One*. 2018. No. 13 (10). DOI: 10.1371/journal.pone.0205417
 20. Pho P. D. *Research Article Abstracts in Applied Linguistics and Educational Technology: a Study of Linguistic Realizations of Rhetorical Structure and Authorial Stance* / P.D. Pho. 2008. Vol. 10. Iss. 2. Pp. 231-250. DOI: 10.1177/1461445607087010
 21. Sinclair J. *Corpus, Concordance, Collocation* / J. Sinclair. Oxford, 1991. 179 p.
 22. Stubbs M. *British traditions in text analysis: from Firth to Sinclair* / M. Stubbs // *Text and Technology* / M. Baker, F. Francis, E. Tognini-Bonelli (eds.). Amsterdam: John Benjamins, 1993. Pp. 1-36.
 23. Swales J. *Genre Analysis: English in Academic and Research Settings* / J. Swales. Cambridge, UK: Cambridge University Press, 1990. 260 pp.

Saida B. KHABIBULLINA¹

Olga B. ULYANOVA²

UDC 82-193

CORPUS ANALYSIS OF REPORTING VERBS IN ABSTRACTS TO RESEARCH ARTICLES

¹ Senior Lecturer, Department of English Philology
and Translation, University of Tyumen
s.b.khabibullina@utmn.ru

² Cand. Sci. (Phylol.), Associate Professor,
Department of English language,
University of Tyumen
o.b.ulyanova@utmn.ru

Abstract

The authors of this article employ the methods of corpus linguistics to study the semantics of general scientific verbs of the lexical-semantic group of reporting in order to study the semantic organization and thematic ordering of this group of English-language predicates in abstracts. The categorical taxonomic meaning of reporting verbs provides an appropriate perception of information when compressing the main text of a research article. Studies that exist in this area comprise the analyses of the rhetorical structure or linguo-cognitive organization of research articles abstracts in various subject areas. Paradigmatics and syntagmatics of lexical units in general and predicates in particular remain not fully understood within the framework of abstracts.

Consequently, the relevance of the subject of the study, namely verbs of reporting in abstracts, is due to the objective need to perform the communicative task of creating or translating a research article abstract mainly from Russian into English. At the same time, the non-English academic community needs access to authentic research, the understanding of which occurs mainly basing on proposition predicates. Based on the material of the compiled corpus of 500 research article abstracts in the subject field of linguistics, the use of automated quantitative

Citation: Khabibullina S. B., Ulyanova O. B. 2020. "Corpus Analysis of Reporting Verbs in Abstracts to Research Articles". Tyumen State University Herald. Humanities Research. Humanitates, vol. 6, no. 1 (21), pp. 62-75.

DOI: 10.21684/2411-197X-2020-6-1-62-75

and qualitative methods of corpus analysis makes the selection of predicates and forms the lexical-semantic group of reporting with the semantic dominant to show, which reveals the highest frequency of use in abstracts. Along with the nuclear semantics *to show*, the semes: *emergence of knowledge; confirmation of knowledge; clarification of knowledge; accentuation of knowledge; overview of knowledge*, organize the space of the lexical-semantic group of reporting and, therefore, the texts of abstracts. Syntagmatics of the studied verbs is limited to four types of combination models of a verb and a direct object; a verb and a prepositional object; a verb and a subordinate clause; as well as a verb and an infinitive, where the first model is most frequent and the last one is least frequent.

Keywords

Corpus analysis, lexical-semantic group, compatibility, frequency of usage.

DOI: 10.21684/2411-197X-2020-6-1-62-75

REFERENCES

1. Ahmanova G. I., Vilshtain A. M., Glushko M. M. 1989. The Semantic Analysis of Reporting Verbs in the Scientific Text. Mocsow: MSU. [In Russian]
2. Baranov A. N. 2001. Introduction to Applied Linguistics. Moscow: Editorial URSS. [In Russian]
3. Bogdanova S. Yu. 2012. “Corpus Methodology Opportunities in Linguistic Tasks”. *Vestnik IGLU*, no. 2yu (18), pp. 47-50. [In Russian]
4. Zakharov V. P. 2003. “Internet Search Engines as a Tool for Linguistic Research”. *Russkiy Yazik v Internete*. Kazan: Otechestvo, pp. 48-59. [In Russian]
5. Zakharov V. P., Bogdanova S. Yu. 2013. Corpus Linguistics. StP: StPSU. RIO. [In Russian]
6. Ibragimova V. L. 2014. “Verb in Semantic Lexicon Environment”. *Vestnik Bashkirskogo Universiteta*, vol. 19, no. 2, pp. 500-505. [In Russian]
7. Lavrentiev A. M. 2004. “Corpus Linguistics: ideology, methods, technology”. *Sibirskiy filologicheskiy журнал*, no. 3-4, pp. 121-134. [In Russian]
8. Paducheva E. V. 2009. “Lexical Aspect and Verb Classification according to Maslow-Vendler”. *Voprosy Yazikoznaniya*, no. 6, pp. 3-20. [In Russian]
9. Ryabtseva N. K. 2014. “Construction grammar, combinatory dictionaries, modern information technologies and foreign language learning”. *Lingvistika i metodika prepodavaniya inostrannyh yazikov*. Moscow: Institut Yazikoznaniya RAN, no. 6, pp. 214-247. Accessed on 13 February 2020.
URL: http://www.iling-ran.ru/library/sborniki/for_lang/2014_06/9.pdf. [In Russian]
10. Biber D., Conrad S., Reppen R. 1998. “Corpus Linguistics: Investigating Language Structure and Use”. Cambridge: Cambridge University Press.
11. Cambridge Dictionary Online URL: <https://dictionary.cambridge.org>.
12. Hu G., Cao F. 2011. “Hedging and Boosting in Abstracts of Applied Linguistics Articles: A Comparative Study of English- and Chinese-medium Journals”. *Journal of Pragmatics*, vol. 43, i. 11, pp. 2795-2809.
URL: <https://www.sciencedirect.com/science/article/abs/pii/S0378216611001135>

13. Hyland K. 2000. Disciplinary Discourses: Social Interactions in Academic Writing. London, UK: Longman.
14. Kingsbury P., Palmer M. 2002. From TreeBank to PropBank. University of Pennsylvania. URL: <http://www.lrec-conf.org/proceedings/lrec2002/pdf/283.pdf>
15. Kipper K. 2005. VerbNet: a Broad-coverage, Comprehensive Verb Lexicon. University of Pennsylvania. Kipper Schuler K. VerbNet: a Broad-coverage, Comprehensive Verb Lexicon: a dissertation in Computer and Information Science / K. Kipper Schuler. University of Pennsylvania, 2005. URL: <http://verbs.colorado.edu/~kipper/Papers/dissertation.pdf>
16. Kipper K., Kornonen A., Ryant N., Palmer M. 2008. "A Large-Scale Classification of English Verbs". *Language Resources and Evaluation Journal*, no. 42, pp. 21-40.
17. Leech G. 1992. "Corpora and Theories of Linguistic Performance". In J. Svartvik (ed.). *Directions in Corpus Linguistics: proceedings of the Nobel Symposium 82*. Berlin; New York: Mouton de Gruyter. *Trends in Linguistics*, vol. 65, pp. 105-122.
18. Levin B. 1993. "English Verb Classes and Alternations: a Preliminary Investigation". Chicago: The University of Chicago Press. 366 p.
19. Ngai S. B. C., Singh R. G., Koon A. C. 2018. "A Discourse Analysis of the Macro-structure, Metadiscoursal and Microdiscoursal Features in the Abstracts of Research Articles across Multiple Science Disciplines". *PLoS One*, no. 13(10). DOI: 10.1371/journal.pone.0205417
20. Pho P. D. 2008. "Research Article Abstracts in Applied Linguistics and Educational Technology: a Study of Linguistic Realizations of Rhetorical Structure and Authorial Stance". Vol. 10, i. 2, pp. 231-250. DOI: 10.1177/1461445607087010
21. Sinclair J. 1991. *Corpus. Concordance. Collocation*. Oxford. 179 p.
22. Stubbs M. 1993. "British Traditions in Text Analysis: From Firth to Sinclair". In M. Baker, F. Francis and E. Tognini-Bonelli (eds.). *Text and Technology: In honour of John Sinclair*. Amsterdam: John Benjamins Publishing, pp. 1-36.
23. Swales J. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge, UK: Cambridge University Press.