

СИНТАГМАТИЧЕСКАЯ И ПАРАДИГМАТИЧЕСКАЯ ВЕРОЯТНОСТЬ ПРИ ИНТЕРПРЕТАЦИИ КОРПУСНЫХ ДАННЫХ (на материале анализа словосочетаний)

Оксана Тарасовна Косаренко¹, Сергей Викторович Косаренко²

Воронежский государственный педагогический университет^{1,2}
Воронеж, Россия

¹Кандидат филологических наук, доцент кафедры русского языка, современной русской и зарубежной литературы, ORCID ID: 0000-0002-5657-7023, тел.: (473) 2747-008, e-mail: o.gordey@mail.ru

²Кандидат филологических наук, доцент кафедры теории, истории и методики преподавания русского языка и литературы, ORCID ID: 0000-0003-1338-7534, тел.: (473) 2747-008, e-mail: serg.v.kosarenko@gmail.com

Аннотация. В статье рассматриваются такие квалификационные признаки словосочетаний, как статистическая устойчивость и регулярность употребления, верифицируемые по метаданным Национального корпуса русского языка. Полученные числовые показатели используются при расчете синтагматической и парадигматической вероятности синтаксической единицы как целого образования, так и ее отдельных форм. Различия в вероятностях являются индивидуальной характеристикой словосочетания, выделенного из ряда номинативных единиц одного семантического поля. В отклонении соответствующих вероятностей (независимой и условной) проявляется языковая закономерность, связанная с необходимостью маркировать словоформы и формы словосочетаний как явление заурядное или необычное. На материале семантического поля «Авторитет» впервые экспериментально проверен инструментальный НКРЯ для установления соотношения способов выражения значений коллокаций. Корпусные методы дают информацию о формальной стороне языковых знаков, их комбинаторике и статистике, но представляют результаты алгоритмов поиска по-разному, что следует учитывать при установлении корреляций и закономерностей функционирования речевых единиц. Более того, однозначность результатов корпусной обработки текстов относительна, так как полученные метаданные меняются в зависимости от учета тех или иных факторов и тождественны средствам, позволяющим представить языковую действительность.

Ключевые слова: словосочетание, n-грамма, речевая единица, валидность, метаданные, частотность, система речи, вероятность, синтагматика, парадигматика.

Для цитирования: Косаренко О.Т., Косаренко С.В. Синтагматическая и парадигматическая вероятность при интерпретации корпусных данных (на материале анализа словосочетаний) // Известия Воронежского государственного педагогического университета. 2022. № 3. С. 155–159. DOI: 10.47438/2309-7078_2022_3_155

Введение

Словарь словосочетаний представляет собой метаязыковые представления об устойчивых отношениях между компонентами данной синтагматической структуры [1]. Как известно, именно синтаксическая зависимость одного из компонентов от другого объединяет сочетание ЛЕ в единое целое. Вторым обязательным условием образования новой сложной единицы является статистическая устойчивость в речевом употреблении. Для оценки числовых показателей встречаемости в текстах (речевой вероятности коллокаций и словоформ) все чаще используют НКРЯ [2; 3; 4], который предлагает в зависимости от запроса исследователя многовариантные результаты обработки массива данных. В связи с этим возникает потребность проверить ва-

лидность корпусных методик при изучении парадигматических и синтагматических координат речевых единиц с учетом их вероятностных дифференциальных признаков.

Результаты

Чтобы определить, в какой мере для получения объективных, полных и точных сведений можно опереться на данный информационный ресурс, рассмотрим наиболее полный по составу словарь словосочетаний, приведенный на сайте «Карта слов и выражений русского языка» [5], в частности 75 атрибутивных сочетаний имен прилагательных со словом *авторитет*. Вследствие того, что неизвестны основания (эмпирические, статистические), по которым были включены те или иные сочетания в состав относительно устойчивых и регулярно употреблявшихся в прошлом или по-прежнему активных словосочетаний, то в общетеоретическом и

практическом плане приобретает актуальность вопрос о принципах отбора синтаксических единиц в состав словника.

Допустим, нас интересует контраст между парадигматической и синтагматической условной вероятностью [6] атрибутивного сочетания *большой авторитет*. Так, конкретная форма числа и падежа для каждого компонента составляет 1/12 в случае независимой вероятности, а отклонения в зависимости от условий употребления и поискового запроса, которые мы намерены проверить, будут учтены при расчете условной вероятности.

1. Задав в основном поиске корпуса расстояние 1, на выходе получаем 223 вхождения словоформ. Форма т.п. ед.ч. (т. ед.) *больш-им авторитет-ом* распределена постранично в соотношениях 30/50 (т.е. к общему количеству всех словоформ), 35/50, 34/50, 30/50, 20/23. Следовательно, 159 падежных форм т. ед. в предложениях составляют значительную часть – 0,71, а синтагматическая вероятность согласования в данном числе и падеже во много раз превосходит парадигматическую вероятность. Продолжим поиск по коллокациям, например 3- и 5-грамм, собранным в кластеры по 100 наиболее частотных последовательностей: триграммы дают соотношение 39/100, а 5-граммы – 86/100. Возникает вопрос, должно ли давать такой разброс сегментирование контактно расположенных членов синтагмы, не разъединенных ни семантически, ни синтаксически. Вообще, будут ли релевантны характеристикам словосочетания *большим авторитетом* эти различающиеся показатели (0,71, или 0,39, или 0,86), или некоторые из них следует рассматривать как промежуточные данные, в отношении которых позднее будет осуществлена должная коррекция? Предположительно, предлагаемые информационной системой алгоритмы обработки дают ответ на запросы в разном приближении к научной истине, при этом разработчики не разъясняют пользователям, какие именно способы дают более достоверную картину языковой действительности.

2. Сопоставим установленные сведения с другими показаниями корпуса, необходимыми для расчета независимой синтагматической вероятности отдельно для прилагательного *большой* и отдельно для существительного *авторитет* (А-т).

2.1. Сначала изменим условия в направлении абстрагирования от конкретного лексического наполнения позиции определяемого слова. В этой ситуации прилагательное *большой* в морфологической форме т. ед. в системе речи имеет 326250 вхождений при условии наличия рядом с ним на расстоянии 1 какого-либо имени существительного. Судя по содержанию списка 3-грамм, *авторитет* не входит в валентность прилагательного.

Функциональный вес формы т. ед. равен 0,18 (18/100) в сочетаниях с такими именами существительными: по модели *с + т. ед. прилагательного + Т. ед. существительного* – *с большим трудом/удовольствием/успехом/интересом/вниманием/количеством/уважением/опозданием/чувством/вкусом/достоинством/любопытством/числом/трудом/аппетитом/-подъемом/нетерпением*. Единичным примером представлена модель *за + т. ед. прилагательного + т. ед. существительного* – *за больш-им стол-ом*.

В лексический спектр кластера 5-грамм «большой» слово *авторитет* не входит. Набор валентностей и синтаксических конструкций подобен кла-

стеру триграмм, но с меньшим лексическим разнообразием имен: *трудом/уважением/удовольствием/интересом/достоинством*. Всего функциональный вес формы Т. ед. равен 0,17 (17/100). Как видим, в количественном и в меньшей степени в качественном отношении показатели близки.

Существительное *авторитет* в связке с каким-либо прилагательным имеет существенно меньше вхождений – 1388. В кластере 3-грамм «авторитет» форм Т. ед. – 8/100, из них только 1 атрибутивное сочетание – *нравственн-ым а-том*. Востребованы иные коллокации с другими именами существительными: *а-том ... власти/истины/ ученых/держав*.

В кластере 5-грамм «авторитет» встречаемость форм т. ед. несколько выше – 11/100, прилаг. *большой* отмечено только в 1 примере и не в связи с формой т. ед.: *а-т больший нежели ...*

2.2. Далее рассмотрим валентности прилагательного *большой* (Б-й) вне зависимости от условий необходимости располагаться рядом с именем существительным, соответственно синтагматическая вероятность атрибутивных сочетаний с данной ЛЕ будет относительно независимой. Прилагательное имеет 641417 вхождений, и, так как анализ среди многостраничных цитаций затруднен, воспользуемся только n-граммами. В кластере 3-грамм форм т. ед. *больш-им* найдено 9/100, все конструкции используют предлог *с*: *трудом/удовольствием/интересом/успехом/вниманием/количеством/уважением*. Лексема *авторитет* отсутствует в перечне. В кластере 5-грамм форм Т. ед. всего 2/100: *между б-им и указательным пальцами; с б-им или меньшим успехом*.

Вне заданной конструкции [прилагательное + существительное], в отрыве от прилагательного встречаемость в текстах существительного *авторитет* увеличивается до 11979 вхождений. При этом в кластере 3-грамм форма т. ед. *авторитет-ом* отмечена 28/100 раз, среди частотных коллокаций употребительны также сочетания с согласованными формами прилагательного *большой*. А в кластере 5-грамм частотность форм т. ед. резко снижается до 17/100 с глаголом – *пользоваться б-им а-ом*. Отсюда следует вывод: для прилагательного валентность к слову *большой* + *авторитет* является периферийной, тогда как для существительного *авторитет* данная валентность входит в ядро.

Строение словосочетаний можно охарактеризовать путем перечисления словоформ, которые составляют речевой образец, через определение наиболее употребительных форм, парадигматическая вероятность которых возрастает/уменьшается в определенных конструкциях. Оставаясь в пределах плана выражения, можно утверждать, что морфологическая форма т. ед. не востребована в комбинаторике каждого из компонентов (для 1-го распределение частот по n-граммам от 0,17–0,18 или от 0,02–0,09; для 2-го от 0,08–0,11 или от 0,17–0,28), тем контрастнее ситуативная синтагматика словосочетания (0,39–0,86), которая своей формой указывает на семантику целого сочетания, связанную с формами т. ед. Тем самым в очередной раз подтверждается основополагающий принцип: сумма лексических и грамматических значений элементов конфигурации не объясняет качественно-количественную характеристику сложного единства.

Обследованное словосочетание (с) *больш-им + авторитет-ом* соответствует типологическим чертам флективного русского языка, удовлетворяя

практически всем критериям, имеющим эксплицитно формальный характер: длина фразы с включенной проклитикой – 19 графем, т.е. ближе к верхнему пределу нормы, актуальному для художественного подъязыка (средневзвешенные данные по языку 12,2–20,3); средний размер словоизменительного аффикса – 2 (-им/-ом), как и в языке в целом; инверсионность – типичная (конфигурация с преобладанием прямых связей над обратными), позиция определения по отношению к определяемому – слева; компактность синтаксиса – мера преобладания ближних связей над дальними (ср. в среднем по языку менее 50 %, т.е. преобладание «слабое», а в рассматриваемом сочетании выражено «сильное» преобладание – более 50 %).

Далее определим распределение функциональной нагрузки на сочетания с однокоренными синонимами-прилагательными *наибольший* и *большой*.

1. Словосочетание «*наибольший авторитет*» – доступно для анализа 15 документов и 15 вхождений в сочетании (чаще всего) с глаголом *пользоваться* и реже с глаголами, управляющими т.п., – *обладать/считать/являться* + превосходн. ст. *наибольшим + а-том*.

2. Словосочетание «*в большом авторитете*». В найденных 203 документах по запросу *большой + а-т* из 223 вхождений приведен только 1 пример употребления предложного субстантивного словосочетания с п.п. существительного и зависимого от него прилагательного: *он был у ... (кого-л.) в большом а-те*.

Перепады между значениями условной вероятности на фоне независимой, интуитивно известные носителю общего фонда знаний данного языкового

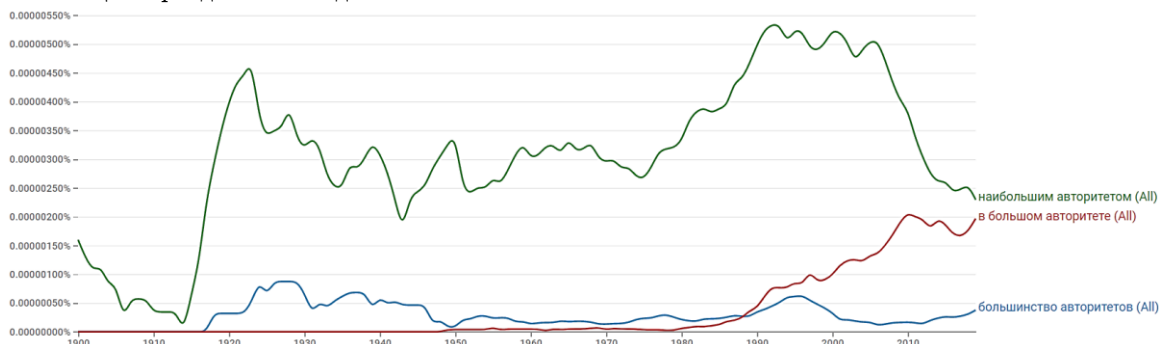


Рис. 1 – Диахронические изменения с 1900 по 2019 г. частот встречаемости сочетаний *наибольшим авторитетом*, *в большом авторитете*, *большинство авторитетов* по данным Google Ngram Viewer

Поиск по изданиям книг за указанный период показывает статистическую устойчивость словосочетаний *наибольшим а-том* и *в больш-ом а-те*, частотность которых выравнивается в 2010–2019 гг. (рис. 1). Субстантивное словосочетание *большинство а-тов* демонстрирует стабильность в употребительности, хотя в прошлом (с 1915 г. до 1990 г.) оно было более востребовано, чем словосочетание *в больш-ом а-те*. В словаре не зафиксировано, например, адъективно-субстантивное словосочетание в форме И.-В. ед. **большой авторитет*, которое по частоте встречаемости в текстах превосходит каждое из трех тестируемых сочетаний. Кроме того, прилагательное *большой* является устойчивой ассоциативной реакцией на стимул АВТОРИТЕТ, что отмечено в словарях ассоциаций в ходе опросов 1986–1997 гг. (РАС [7]) и 2008–2020 гг. (СИБАС [8; 9; 10]).

коллектива, маркируют те или иные словосочетания по стилистическим параметрам. Проверим это утверждение, сравнив отнесенность к типу текста названных сочетаний. Если конструкция *большой + а-т* во всех найденных формах падежа и числа используется преимущественно в мемуарной литературе (66 вхождений, или 29,60 %) и дневниковых записях (47 вхождений, или 21,08 %), а также в статьях (22, или 9,87 %), то сочетание *наибольшим + а-том* выделяется на этом фоне закрепленностью в текстах статей (7, или 46,67 %) и монографий (2 вхождения, или 13,33 %), причем публикации относятся в большей степени к сфере публицистики (ср.: 8, или 53,33 %, для *большой а-т* и 4, или 26,67 % для *наибольшим а-том*), чем к учебно-научной сфере.

Результаты чувствительны к заданному в меню поиска расстоянию между компонентами. Так, Корпус не дает примеры контактного расположения ЛЕ с беспредложным субстантивным словосочетанием *большинство авторитетов* с зависимым существительным в р.п. А на расстоянии от -3 ... 3 в Корпусе найдено 7 документов, 7 вхождений, из которых только 2 коллокации образуют сложную синтаксическую единицу: *большинство криминальных/научных а-тов*. Есть некоторая корреляция между позицией компонентов в предложении и частотностью данной валентности: поиск по n-граммам (кластерам 3-грамм и 5-грамм) убеждает, что в высокочастотную валентность существительного *большинство* не входит зависимое существительное *авторитет*.

Место каждого словосочетания в системе речи и конкретно в ряду конкурирующих сочетаний определено его местом в семантическом пространстве и в ряду грамматических значений (в плане смыслового устройства), а в плане формы – на пресечении парадигматической и синтагматической вероятности, на координатах условной и независимой вероятности образования словосочетания. Поэтому частотность употребления конструкции целесообразно верифицировать с учетом частотности отдельных форм, т.е. условной парадигматической вероятности. Так, словесная реакция *непререкаемый* на уровне предложения является левой валентностью ближайшего определяющего слова *авторитет*, а сочетание имеет 150 вхождений, сконцентрированных в 139 документах. Парадигма падежных словоформ не является полной: ед.ч. и.-в. -ый ... -ш – 47, р. -ого ... -а – 7, д. -ому ... -у – 5, т. -ым ... -ом – 81, непререкаем-ш ... авторитет-ш – 1; мн.ч. -ые ... -ы – 2, -ых ... -

ов – 3, -ыми ... -ами – 3. Шестипадежная система единственного числа представлена в самых употребительных формах и.-в.п. – 47 и т.п. – 81.

Сходное морфологическое строение коррелирует со средней частотностью использования в речи сравниваемых словосочетаний. Рассмотрим подробнее функциональные различия, определяющие выбор между однокоренными синонимами *беспорный/неоспоримый а-т*. Из 19 вхождений 18 образуют синтаксически связанные словосочетания *неоспоримый а-т* в разных падежных формах компонентов (ед.ч. -ый ... -ш – 3; -ого ... -а – 4; -ым ... -ом – 9; мн.ч. -ые ... -ы – 2). В 1949 г. максимум был равен 0,225, затем происходит снижение активности, а с 1975 г. начинается медленный рост встречаемости до максимума в 2003 г. (0,130) с последующим постепенным снижением до 0 в 2015 г. *Беспорный а-т* имеет 18 вхождений с более полной парадигмой форм (ед.ч. -ый ... -ш – 6, -ого ... -а – 2;

-ым ... -ом – 8; мн.ч. -ые ... -ы – 1, -ыми ... -ами – 1). Вершина популярности словосочетания пришла на 1950-е годы (в 1954 г. был достигнут максимум – 0,342), в 2006 г. 0,117, потом кривая опустилась до 0 в 2015 г. Как видим из сравнения, популярность сочетаний с *беспорный* встречается за счет большей употребительности в текстах, относящихся к советской эпохе. Но сфера функционирования словосочетания *беспорный а-т* шире (6, включая помимо публицистики – 45 % и художественных текстов – 15 % еще учебно-научную, бытовую, церковно-богословскую сферу по – 10–15 %), чем у словосочетания *неоспоримый а-т* (3, из которых больше 94 % приходится на публицистику с художественной литературой). Общим морфологическим признаком обоих словосочетаний является наибольшая активность формы т.п. ед.ч. На графике их кривые, выражая близкие числовые показатели использования в книгах, переплетаются (рис. 2).

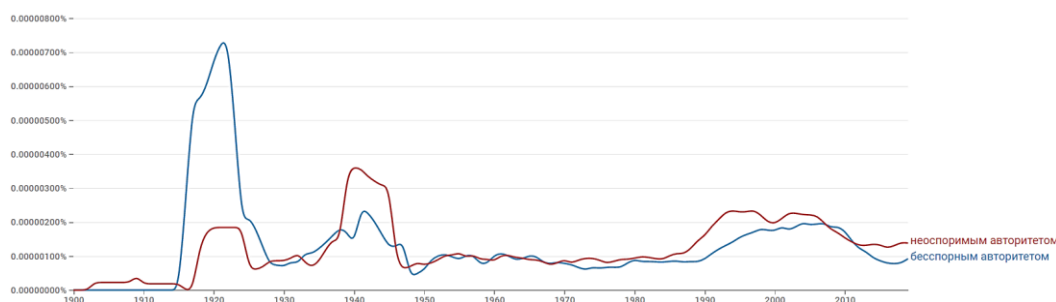


Рис. 2 – Диахронические изменения с 1900 по 2019 г. частот встречаемости сочетаний *беспорным авторитетом*, *неоспоримым авторитетом* по данным Google Ngram Viewer

Выводы

Воздействие априорно признаваемых экстралингвистических факторов, подталкивающих к изменениям в системе речи на уровне словосочетаний и конфигурации исследуемой ЛСГ, опосредованно внутриязыковыми соотношениями парадигматических/синтагматических, условных/независимых вероятностей события – возникновения словосочетания в совокупности его словоформ и регулярного его воспроизведения в предложении. Резкие перепады частотных показателей служат средством выделения словосочетания по способу функционирования в речи, его стилистической маркировки по сфере употребления.

Валидация инструментов НКРЯ применительно к задаче определения функциональной нагрузки форм словосочетаний показала широту возможностей алгоритмов поиска и его гибкую настройку по заданным характеристикам, однако исходом программной обработки являются не объективные данные, а промежуточные метаязыковые описания, которые требуют дальнейшей лингвистической ин-

терпретации. Неоднозначность результатов адекватна сложности находящегося в движении объекта – языковой действительности, для метаязыкового описания которого приняты вероятностные дифференциальные признаки языковой единицы. Имеет место многоаспектная репрезентация выданного материала (ср., например, содержание n-gram), не только стимулирующая дальнейший поиск, но и затрудняющая интерпретацию. Поэтому эти особенности необходимо учитывать в организации самостоятельной поисковой исследовательской работы при постановке и решении проблемных задач с применением корпусных методов. Для составления «Словаря словосочетаний» необходимо провести корректировку в отношении доказательной базы и верификации методов включения или исключения из него конструкций, что потребует дополнительно изучения.

Конфликт интересов

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Библиографический список

1. Всеволодова М.В. Словосочетание в новой парадигме грамматики // Вестник Московского университета. Сер. 9: Филология. 2012. № 1. С. 7–40.
2. НКРЯ – Национальный корпус русского языка. URL: <http://ruscorpora.ru> (дата обращения: 05.07.2022).
3. НОССРЯ – Новый объяснительный словарь синонимов русского языка / под общим руководством акад. Ю.Д. Апресяна. 2-е изд., испр. и доп. М. : Языки славянской культуры, 2003. 1488 с.
4. Ляшевская О.Н. Корпусные инструменты в грамматических исследованиях русского языка. М. : Рукописные памятники Древней Руси, 2016. 518 с.
5. КартаСлов.ру: онлайн-карта слов и выражений русского языка. URL: <https://kartaslov.ru/> (дата обращения: 05.07.2022).
6. Андреев Н.Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении / АН СССР. Ин-т языкознания. Ленинград : Наука. Ленингр. отд-е, 1967. 403 с.

7. РАС – Русский ассоциативный словарь // TESAUROS.RU: Информационная система когнитивных экспериментов. URL: <http://www.thesaurus.ru/dict/> (дата обращения: 05.07.2022).
8. СИБАС: Русская региональная ассоциативная база данных – Сибирь и Дальний Восток / авт.-сост.: И.В. Шапошникова, А.А. Романенко. URL: <http://adictru.nsu.ru> (дата обращения: 05.07.2022).
9. Косаренко О.Т., Косаренко С.В. Структурные отношения в ассоциативно-вербальной сети «Авторитет» в современном русском языке // Известия Воронежского государственного педагогического университета. 2021. № 3. С. 145–149. DOI: 10.47438/2309-7078_2021_3_145.
10. Караулов Ю.Н. Ассоциативная грамматика русского языка. Изд. 2-е. М. : Либроком, 2010. 330 с.

References

1. Vsevolodova, M.V. (2012) Slovosochetanie v novoi paradigme grammatiki [Word-combination in the new paradigm of grammar]. *Vestnik Moskovskogo universiteta. Philology.* (1), 7–40. (in Russian)
2. NKRYa – *Natsional'nyi korpus russkogo yazyka* [National corpus of the Russian language]. Available from: <http://ruscorp.ru> [Accessed 05.07.2022].
3. Apresyan, Yu.D. *NOSSRYa – Novyi ob'yasnitel'nyi slovar' sinonimov russkogo yazyka* (2003) [The New Explanatory Dictionary of Russian Synonyms]. Moscow, Yazyki slavyanskoi kul'tury publ. 1488 p. (in Russian)
4. Lyashevskaya, O.N. (2016) *Korpusnye instrumenty v grammaticheskikh issledovaniyakh russkogo yazyka* [Corpus Instruments for Russian Grammar Studies]. Moscow, Rukopisnye pamyatniki Drevnei Rusi publ. 518 p. (in Russian)
5. KARTASLOV.RU: onlain-karta slov i vyrazhenii russkogo yazyka [online map of Russian words and expressions]. Available from: <https://kartaslov.ru/сочетаемость-слова/авторитет> [Accessed 05.07.2022].
6. Andreev, N.D. (1967) *Statistiko-kombinatornye metody v teoreticheskom i prikladnom yazykovedenii* [Statistical combinatorial methods in theoretical and applied linguistics]. Leningrad, Nauka publ. 403 p. (in Russian)
7. RAS – *Russkii assotsiativnyi slovar'* [Russian associative dictionary]. *TESAUROS.RU: Informatsionnaya sistema kognitivnykh eksperimentov*. Available from: <http://www.thesaurus.ru/dict/> [Accessed 05.07.2022].
8. Shaposhnikova, I.V., Romanenko, A.A. *SIBAS – Russkaya regional'naya assotsiativnaya baza dannykh – Sibir' i Dal'nii Vostok* [Russian regional associative database (Siberia and Far East)]. Available from: <http://adictru.nsu.ru> [Accessed 05.07.2022].
9. Kosarenko, O.T., Kosarenko, S.V. (2021) *Strukturnye otnosheniya v assotsiativno-verbal'noi seti «Avtoriteta» v sovremennom russkom yazyke* [Structural relations in the verbal associative network “Avtoriteta” in modern Russian language]. *Izvestiya Voronezhskogo gosudarstvennogo pedagogicheskogo universiteta.* 3, 145–149. (in Russian). Doi:10.47438/2309-7078_2021_3_145
10. Karaulov, Yu.N. (2010) *Assotsiativnaya grammatika russkogo yazyka* [Associative grammar of the Russian language]. Moscow, Librokom publ. 330 p. (in Russian)

Поступила в редакцию 26.07.2022
Подписана в печать 27.09.2022

Original article

UDC 81-112.4

DOI 10.47438/2309-7078_2022_3_155

SYNTAGMATIC AND PARADIGMATIC PROBABILITY IN THE INTERPRETATION OF CORPUS DATA

Oksana T. Kosarenko¹, Sergej V. Kosarenko²

*Voronezh State Pedagogical University^{1, 2}
Voronezh, Russia*

¹*Cand. Philolog. Sci., Docent of the Department of Russian Language, Modern Russian and Foreign Literature, ORCID ID: 0000-0002-5657-7023, тел.: (473) 2747-008, e-mail: o.gordey@mail.ru*

²*Cand. Philolog. Sci., Docent of the Department of Theory, History and Methodology of Teaching Russian Language and Literature, ORCID ID: 0000-0003-1338-7534, тел.: (473) 2747-008, e-mail: serg.v.kosarenko@gmail.com*

Abstract. The article deals with such qualifying features of word-combinations as statistical stability and regularity of use, verified by the metadata of the National Corpus of the Russian Language. The obtained numerical indicators are used in the calculation of the syntagmatic and paradigmatic probability of syntactic unit of both the whole formation and its individual forms. Differences in probabilities are an individual characteristic of word-combinations selected from a number of nominative units of one semantic field. On the material of the semantic field “Authority”, for the first time, the NCRL toolkit was experimentally tested to establish the correlation of ways to express the meanings of collocations. Corpus methods provide information about the formal side of linguistic signs, their combination and statistics, but present the results of search algorithms in different ways, and this should be taken into account when establishing correlations and regularities in the functioning of speech units.

Key words: word-combination, n-gram, unit of speech, validity, metadata, frequency, speech system, probability, syntagmatics, paradigmatics.

Cite as: Kosarenko, O.T., Kosarenko, S.V. (2022) Syntagmatic and paradigmatic probability in the interpretation of corpus data. *Izvestia Voronezh State Pedagogical University.* (3), 155–159. (In Russ., abstract in Eng.). DOI: 10.47438/2309-7078_2022_3_155

Received 26.07.2022

Accepted 27.09.2022