

О методе комплексного семантического, статистического и психолингвистического анализа многозначности

Валентина Юрьевна Апресян^{1,2} (valentina.apresjan@gmail.com), Юрий Дереникович Апресян^{1,3} (juri.apresjan@gmail.com), Ольга Викторовна Драгой² (odragoy@hse.ru), Борис Леонидович Иомдин^{1,2} (iomdin@ruslang.ru), Анна Кестучё Лауринавичюте² (alaurinavichute@hse.ru), Ирина Борисовна Левонтина¹ (irina.levontina@mail.ru), Константин Александрович Лопухин⁴ (kostia.lopuhin@gmail.com), Анастасия Александровна Лопухина^{1,2} (nastya.lopukhina@gmail.com), Елена Владимировна Урысон¹ (uryson@gmail.com)

¹ Институт русского языка им. В. В. Виноградова РАН

² Национальный исследовательский университет «Высшая школа экономики»

³ Институт проблем передачи информации им. А. А. Харкевича РАН

⁴ Scrapinghub

DOI: 10.31857/S013161170003937-9

аннотация: Цель исследования — мультидисциплинарное изучение феномена полисемии (многозначности) языковых единиц с помощью теоретических, экспериментальных и статистических методов. Хотя полисемии посвящено большое количество работ, это явление ранее не исследовалось комплексно. Коллективом авторов было проведено исследование, которое сочетало элементы словарного описания, статистического анализа, опросов, а также изучение электроэнцефалограмм и движений глаз. Исследование показало, что при развитии полисемии используется большее количество различных семантических сдвигов, помимо хорошо известных метафоры и метонимии. Эти сдвиги составляют сложную иерархическую систему и часто комбинируются друг с другом при образовании новых значений. Наше восприятие значения

как нового связано с когнитивным «расстоянием», которое различается для разных типов сдвигов: так, метафорически образованное значение воспринимается как более далекое от исходного, чем метонимическое значение. Словарное представление значений только отчасти коррелирует с устройством ментального лексикона и с частотностью разных значений. Лексикографическое представление, основанное на семантических принципах, более удобно для восприятия, чем представление, основанное на частоте употребления. В ходе исследования возникли новые вопросы, в частности, различаются ли представления далекой и близкой метонимии в нашем ментальном лексиконе.

КЛЮЧЕВЫЕ СЛОВА: семантика, лексикография, многозначность, семантические сдвиги, частотность, корпус текстов, семантическая близость, ментальный лексикон, экспериментальная лингвистика

для цитирования: Апресян В. Ю., Апресян Ю. Д., Драгой О. В., Иомдин Б. Л., Лауринавичюте А. К., Левонтина И. Б., Лопухин К. А., Лопухина А. А., Урысон Е. В. О методе комплексного семантического, статистического и психолингвистического анализа многозначности // Русская речь. 2019. № 1. С. 8–17. DOI: 10.31857/S013161170003937-9

БЛАГОДАРНОСТИ: Статья написана по результатам работы над грантом РФФ № 16-18-02054 «Исследование русского языкового сознания на основе семантического, статистического и психолингвистического анализа лексической многозначности».

Issues of Modern Russian Language

A Multifaceted Approach to Semantic, Statistical, and Psycholinguistic Analysis of Lexical Polysemy

Valentina Yu. Apresyan^{1,2} (valentina.apresjan@gmail.com), Yury D. Apresyan^{1,3} (juri.apresjan@gmail.com), Olga V. Dragoy² (odragoy@hse.ru), Boris L. Iomdin^{1,2} (iomdin@ruslang.ru), Anna K. Laurinavichyute² (alaurinavichute@hse.ru),

Irina B. Levontina¹ (irina.levontina@mail.ru), Konstantin A. Lopukhin⁴
(kostia.lopuhin@gmail.com), Anastasiya A. Lopukhina^{1,2} (nastya.lopukhina@gmail.com),
Elena V. Uryson¹ (uryson@gmail.com)

¹ Vinogradov Russian Language Institute (Russian Academy of Sciences)

² National Research University «Higher School of Economics»

³ Institute for Information Transmission Problems (Kharkevich Institute) (Russian Academy of Sciences)

⁴ Scrapinghub

ABSTRACT: The article is a cross-disciplinary study of polysemy applying theoretical, experimental and statistical methods. Although polysemy has been a focus of attention in a variety of papers and books, the phenomenon requires a largely complex approach. The team of authors conducted a multifaceted study that incorporated lexicographic descriptions, statistical analysis, speakers' surveys, as well as experimental studies of electroencephalograms and eye movements. The study revealed that along with metaphors and metonymy, which are the most common meaning shifts, the development of polysemy also involves a greater number of other shift types. Taken together, these shifts constitute an elaborate hierarchical system and often combine with each other when new meanings are formed. Whether or not the speaker perceives that a word meaning is new, depends on the cognitive distance, which varies for different types of shifts: specifically, a metaphorically formed meaning is perceived to be further away from the original than a metonymic meaning. The way dictionary entries represent lexical meanings is only partly relevant to the structure of the mental lexicon and to the frequency of occurrence of different meanings. A lexicographic representation based on semantic principles is more readily perceived than one based on meaning frequency. The study has invited new interesting questions, such as whether our mental lexicon discriminates between distant and close metonymy.

KEYWORDS: semantics, lexicography, polysemy, semantic shifts, frequency, corpus of texts, semantic proximity, mental lexicon, experimental linguistics

FOR CITATION: Apresyan V. Yu., Apresyan Yu. D., Dragoy O. V., Iomdin B. L., Laurinavichyute A. K., Levontina I. B., Lopukhin K. A., Lopukhina A. A., Uryson E. V. A Multifaceted Approach to Semantic, Statistical, and Psycholinguistic Analysis of Lexical Polysemy. Russian Speech = Russkaya Rech'. 2019. No. 1. Pp. 8–17. DOI: 10.31857/S013161170003937-9

ACKNOWLEDGEMENTS: This research is supported by a grant from Russian Science Foundation, project No. 16-18-02054 «Study of Russian language consciousness based on semantic, statistical and psycholinguistic analysis of lexical ambiguity».

Одно из фундаментальных и самых захватывающих свойств языковых единиц — это их многозначность (полисемия). Подавляющее большинство слов русского языка, в особенности среди ядерной части лексики, имеет больше одного значения; нередко встречаются слова с более чем десятью значениями. При этом, когда мы сталкиваемся с многозначными словами в тексте или в речи, мы обычно легко понимаем, какое значение многозначного слова имеется в виду. Как такое возможно? Дело в том, что в разных значениях слово имеет разную сочетаемость, т. е. появляется в разных контекстах. Например, когда мы слышим фразу *сок из граната*, то понимаем, что речь идет о значении 'плод', а во фразе *ожерелье из граната* слово *гранат* естественно интерпретируется в значении 'камень'.

Проблема разрешения неоднозначности полисемичных слов уже давно обсуждается в теоретической семантике. Так, в книге Ю. Д. Апресяна «Лексическая семантика» [Апресян 1974] сказано, что для понимания того, как происходит выбор нужной интерпретации, необходим учет законов взаимодействия значений, то есть семантических правил. Наиболее общим является правило согласования значений, которое предполагает максимальную повторяемость смысловых компонентов в пределах предложения.

В книге приводится такой пример. Во фразе *Лошадка везет из леса воз хвороста* мы понимаем, что хворост — это сухие ветки, а во фразе *Бабушка жарит хворост на плите* — что это кулинарное изделие. В первом случае очевидным образом в толкованиях слов *лес* и *хворост* будет повторяться компонент 'дерево', во втором в толкованиях слов *жарить*, *хворост*, *плита* — компонент 'пища' или подобный. При этом у слов *жарить*, *плита* также есть другие значения, и выбор между ними осуществляется по этому же принципу. Если контекст не предоставляет такой возможности, возникает неоднозначное высказывание, как, например, *Хворост еще сырой* (совпадающие компоненты есть при обоих пониманиях слова *хворост*, им соответствуют разные значения слова *сырой*), или абсурдное (*Хворост спит под громкостью*, где у слов слишком мало общего, *Хворост цветет*, где между словами возникает семантическое противоречие). Абсурдные высказывания могут получать интерпретацию в игровом или поэтическом контексте.

Многозначность, с одной стороны, — это способ человеческого мышления, познания и осознания мира. С другой стороны, это языковое явление, которое играет ключевую роль в понимании текстов и речи, в том числе и при автоматической обработке текста.

В лингвистической семантике и лексикографии исследуются регулярные модели многозначности, вырабатываются лингвистические критерии выделения значений. В компьютерной лингвистике, в частности в систе-

мах машинного перевода и автоматической обработки текста, на первый план выходит проблема разрешения многозначности — определения, в каком значении употреблена языковая единица в данном тексте. Ведь без этого невозможно правильно понять этот текст. Отдельное направление анализа многозначности представлено в психо- и нейролингвистике, где, с одной стороны, исследуются проблемы представления многозначности в ментальном лексиконе, а с другой — проблемы доступа к значениям при понимании речи в режиме реального времени, в том числе с точки зрения того, как «работают» при этом нейроны. До сих пор эти подходы к многозначности существовали изолированно — например, результаты психолингвистических и нейролингвистических исследований никак не учитывались в работе над словарными описаниями.

Однако оказывается, что многие теоретические семантические правила, в частности правило согласования значений Ю. Д. Апресяна, находят подтверждение в современной экспериментальной и компьютерной лингвистике. Например, экспериментальные методы демонстрируют, что близкие значения слова ускоряют восприятие друг друга: прямое значение действует как позитивный прайминг (активатор) для метонимического значения, и наоборот; современные автоматические методы разбиения на значения также ориентированы на анализ контекстов.

Поэтому нам представляется оптимальным комплексный анализ полисемии, совмещающий разные подходы к этому сложному явлению и сравнивающий их результаты; мы стремимся к интеграции лексикографических, количественных, психолингвистических и нейролингвистических методов.

На большом языковом материале были выделены и описаны механизмы семантической деривации: метафора, то есть перенос наименования по сходству (*груша* (плод) → *груша* (боксерская)), метонимия, то есть перенос по смежности (*глупый человек* → *глупый вопрос*) [человек не похож на вопрос, но глупый вопрос может быть задан глупым человеком], добавление семантических компонентов (сужение, конкретизация, обогащение значения: *равнобедренный треугольник* → *играть на треугольнике*, к компоненту 'форма' добавляются 'материал' и 'музыкальный инструмент'), утрата компонентов (расширение значения: *человек идет* → *поезд идет* [остаётся компонент 'перемещение', а указание на конкретный способ утрачивается]), замена компонентов (*Я услышал голос* → *Мне голос был*, меняется компонент 'чей'), модализация (добавление оценки: *с трудом запихать в чемодан все вещи* → *Ну куда ты запихал мои ключи?*), гиперболизация (ослабление: *Наш сантехник просто гений*), смягчение (литотизация, эвфемизация: *слова сомнительный, неоднозначный* в значении 'плохой', у слова с более слабым смыслом появляется более сильный смысл 'плохой'), утрата агентивности (*Ребенок разбил* → *Ветер разбил*), перенос

по коннотациям (*разводить свиней* → *Он жуткая свинья, у него всегда такой беспорядок*, по коннотации неаккуратности), перенос по имплицатуре, т. е. явно не выраженному необязательному смыслу (*добрый человек* → *добрый кусок мяса*, скрытый смысл «Хороший — значит большой»), десемантизация («выветривание»: *давать деньги взаймы* → *Давай пойдём в кино*), внешняя мотивация (*левый и правый* в политическом смысле связаны с традиционным расположением мест фракций в английском парламенте), конверсия (*выменять свое новое платье на туфли* — *выменять платье на свои новые туфли*).

Некоторые из типов переносов — например, метафора и метонимия — хорошо известны, однако исследование на массовом языковом материале позволяет расширить представление об их природе. Так, гиперболизация и смягчение (литотизация) обычно рассматриваются как риторические фигуры, используемые для создания художественного или стилистического эффекта. Однако они также представляют собой регулярный механизм развития многозначности.

Некоторые семантические переносы представляют собой более узкие разновидности других. Скажем, модализация — это разновидность добавления компонентов, поскольку к производящему значению добавляется особый семантический компонент — качественная оценка, часто отрицательная. Кроме того, некоторые семантические переносы часто появляются в сочетании с другими. Например, у слова *голос* есть основное значение ‘звучание речи человека’ (*Помню ее голос*), от которого метонимическим переносом образуется значение ‘говорящий человек’ (*Голоса удалялись*). При этом важно, что помимо метонимии здесь есть семантическое наращение: говорящий человек не виден. Сказать *голоса удалялись* нормально, только если обладателей голоса мы не видим.

На материале полисемии ряда прилагательных обнаружилось сближение метонимии, метафоры и прямого значения: а именно, метонимические производные от прямого и метафорического значений могут объединяться в одной лексеме (*безумный поступок* — ‘поступок сумасшедшего’ (метонимия) или ‘как бы являющийся проявлением психической болезни поступок’ (метафора)).

Даже у служебных слов, например предлогов, характер переносов имеет во многих случаях такую же природу. Так, многозначность предлога *до* во многом построена на метафорических и метонимических сдвигах. Метафорический перенос ‘пространство → время’ (*доползти до забора* → *работать до вечера*) — это та же самая метафора, что в выражениях *далекое прошлое* или *длинные выходные*.

Любопытно, что некоторые из переносов приводят к большему семантическому сдвигу, чем другие. Так, метонимия — это менее осязаемый

сдвиг по сравнению с метафорой. Неслучайно люди обычно плохо понимают суть метонимии и саму идею семантической смежности, в то время как понимание сути метафоры обычно затруднений не вызывает. Экспериментальные исследования языковой интуиции носителей языка подтверждают гипотезу о меньшем когнитивном «расстоянии» между метонимическими производными, нежели между метафорическими.

После обработки семантических данных мы провели поведенческое экспериментальное исследование, чтобы понять, как в ментальном лексиконе хранятся значения многозначных слов разных частей речи. Мы использовали задание на классификацию значений, чтобы исследовать, как носители русского языка группируют словосочетания с прямыми, метонимическими и метафорическими значениями слов трех частей речи — существительных, прилагательных и глаголов.

ВОДА

Вся вода испарилась
В докладе много воды
В этом тексте одна вода!
Чемпионат мира по плаванию на короткой воде



создать новую группу



перейти к следующему вопросу

Пример задания на группировку значений существительного *вода*

Example task for grouping meanings of the noun *voda* (water)

Участникам эксперимента было предложено распределить по виртуальным корзинам 6–12 коротких контекстов с одним и тем же словом в разных значениях так, чтобы в каждой корзине это слово выступало в одном и том же значении. Испытуемые могли создавать столько корзин, сколько было контекстов на экране, и группировать контексты так, как им хотелось, например, можно было положить каждый контекст в отдельную корзину или сложить все контексты в одну корзину. В качестве стимулов были выбраны 24 существительных, 12 глаголов и 12 прилагательных. В исследовании приняли участие 2080 добровольцев.

Результаты эксперимента показали, что метонимические значения воспринимаются как близкие к прямым во всех трех классах слов. Однако

у прилагательных метонимические и метафорические значения воспринимаются как более похожие. Это позволяет сделать вывод о том, как хранятся значения многозначных слов: прямые и метонимические значения существительных и глаголов хранятся в одной ментальной репрезентации, а метафорические — в отдельных репрезентациях. Многозначные прилагательные, судя по всему, хранятся иначе: репрезентации прямых значений пересекаются с репрезентациями метонимических значений, а те, в свою очередь, пересекаются с репрезентациями метафорических значений [Лопухина, Лауринавичюте, Драгой 2018].

В ходе эксперимента подтвердилось предположение о том, что лингвисты будут классифицировать значения ближе к теоретическому методу классификации, на который опираются составители словарей. Однако не было найдено никакого подтверждения тому, что нелингвисты пользуются принципиально отличным способом классификации значений — общий рисунок ошибок классификации в двух группах испытуемых похож [Lopukhina, Laurinavichyute, Lopukhin, Dragoy 2018].

Был также проведен эксперимент с использованием методики вызванных потенциалов мозга. Мы измеряли электрическую активность мозга в то время, как испытуемым предъявлялись стимулы-словосочетания (например, *политическая арена — цирковая арена*): о каждом словосочетании нужно было решить, осмысленно оно или нет. Мы предположили, что ответ мозга на второе словосочетание в паре будет разным, в зависимости от того, какое значение у существительного в первом словосочетании: чем дальше друг от друга значения существительного в первом и втором словосочетании, тем выше степень «удивления» участника при восприятии второго словосочетания в паре. Это позволяет определить, какие значения мозг воспринимает как близкие и связанные, а какие — как далекие и несвязанные, подробнее в [Yurchenko, Lopukhina, Dragoy, 2018].

Еще один тип экспериментов исследует механизмы лексического доступа при понимании многозначных слов в предложениях и связан с регистрацией движений глаз при чтении. Так, мы сравнили движения глаз при чтении многозначных слов, выступающих в прямом, метафорическом или метонимическом значениях, например: *груша* — плод (прямое значение), дерево (метонимия) и спортивный снаряд (метафора).

Эксперимент с регистрацией движений глаз при чтении многозначных слов показал, что прямое, метонимическое и метафорическое значения слова читаются одинаково быстро, если поддерживаются предшествующим контекстом. Исключение составляли случаи, когда контекст был настолько общим, что был совместим со всеми значениями, но не поддерживал ни одно из них больше других, тогда неоднозначные слова читались значительно дольше. Таким образом, понимание метонимии

и метафоры не требует дополнительных затрат энергии по сравнению с пониманием прямого значения — при условии, что контекст ограничивает возможные значения.

В статистической (компьютерно-лингвистической) части исследования интересным оказался результат статистического исследования о том, как часто первое значение в словаре оказывается самым частотным в корпусе. Для прилагательных, и особенно для глаголов, первое словарное значение только в половине случаев является наиболее употребительным в современном языке, что доказывает необходимость отражения в словаре информации о частотности значения слова, а с другой стороны, показывает, насколько необходимы данные о частотности значений при вероятностном разрешении многозначности в автоматической обработке текста.

В статистических и психолингвистических исследованиях многозначности активно используются данные существующих словарей. Результаты этих исследований, в свою очередь, могут быть очень полезны в лексикографической работе, особенно с точки зрения повышения удобства словаря для пользователя. Так, были проведены эксперименты, задачей которых было установить, какая подача значений сильно многозначного слова предпочтительна для пользователя. Группам испытуемых были предложены описания слов, имеющих большое количество значений (*вынести* и *вырасти*). Эти описания были упорядочены по частотности значений или в порядке семантического развертывания от прямых значений к переносным, исследовалось, сколько значений и в каком порядке испытуемые потом могут вспомнить. Оказалось, что порядок практически не влияет на количество вспомненных значений, но сам семантический порядок запоминается лучше, чем частотный.

Кроме того, выяснилось, что испытуемые, не получившие никаких материалов, привели некоторые новые, в том числе сленговые, значения, не отраженные в «Активном словаре русского языка» (об Активном словаре см. [Апресян (*ред.*) 2010]), например, весьма частотным оказалось словосочетание *вынести мозг*. Опора на корпус текстов при составлении современных словарей необходима, но недостаточна: новые значения слова, особенно стилистически ярко окрашенные, могут выявляться лишь путем языкового эксперимента.

Интегрированное исследование многозначности пока в самом начале. Однако уже очевидно, что оно позволяет получить объемное описание этого феномена, уточнить семантические описания единиц языка, углубить наши представления о том, как происходит хранение лингвистической информации и ее обработка в мозгу человека, усовершенствовать методы автоматической обработки текста и сделать словари более удобными для пользователя.

Литература

- Апресян Ю. Д. Экспериментальное исследование семантики русского глагола. М.: Наука, 1967. 256 с.
- Апресян Ю. Д. Лексическая семантика. Синонимические средства языка. М.: Наука, 1974. 368 с.
- Апресян Ю. Д. (ред.) Проспект активного словаря русского языка (отв. ред. акад. Ю. Д. Апресян). М.: Языки славянских культур, 2010. 784 с.
- Лопухина А. А., Лауринавичюте А. К., Драгой О. В. Как в ментальном лексиконе хранятся многозначные слова разных частей речи? // В кн.: Восьмая международная конференция по когнитивной науке: Тезисы докладов. Светлогорск, 18–21 октября 2018 г. М.: Институт психологии РАН, 2018. С. 644–646.
- Lopukhina A., Laurinavichyute A., Lopukhin K., Dragoy O. The Mental Representation of Polysemy across Word Classes. In: *Frontiers in Psychology*. 2018. Vol. 9. P. 1–16.
- Yurchenko A., Lopukhina A., Dragoy O. Meaning relatedness in polysemous and homonymous words: an ERP study in Russian. Working papers by the Basic Research Program. Series WP BRP 67/LNG/2018 "Linguistics / LNG". 2018.
-

References

- Apresyan Yu. D. *Eksperimental'noe issledovanie semantiki russkogo glagola*. [Investigational study of the semantics of the Russian verb]. Moscow, Nauka Publ., 1967. 256 p.
- Apresyan Yu. D. *Leksicheskaya semantika. Sinonimicheskie sredstva yazyka* [Lexical semantics. Synonymous language means]. Moscow, Nauka Publ., 1974. 368 p.
- Apresyan Yu. D. (ed.) *Prospekt aktivnogo slovarya russkogo yazyka (otv. red. akad. Yu. D. Apresyan)*. Moscow, Yazyki slavyanskikh kul'tur Publ., 2010. 784 p.
- Lopukhina A. A., Laurinavichyute A. K., Dragoy O. V. [How are polysemantic words of different parts of speech kept in the mental lexicon?]. *Vos'maya mezhdunarodnaya konferentsiya po kognitivnoi nauke: Tezisy dokladov. Svetlogorsk, 18–21 oktyabrya 2018 g.* [8th international conference on cognitive science. Abstracts. Svetlogorsk, October 18–21, 2018]. Moscow: Institute of Psychology of RAS, 2018, pp. 644–646. (In Russ.)
- Lopukhina A., Laurinavichyute A., Lopukhin K., Dragoy O. The Mental Representation of Polysemy across Word Classes. In: *Frontiers in Psychology*. 2018. Vol. 9. Pp. 1–16.
- Yurchenko A., Lopukhina A., Dragoy O. Meaning relatedness in polysemous and homonymous words: an ERP study in Russian. Working papers by the Basic Research Program. Series WP BRP 67/LNG/2018 "Linguistics / LNG". 2018.