

Вестник Костромского государственного университета. 2025. Т. 31, № 3. С. 190–194. ISSN 1998-0817

Vestnik of Kostroma State University, 2025, vol. 31, no. 3, pp. 190–194. ISSN 1998-0817

Научная статья

5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика.

УДК 81.161.1'373

EDN PLCLOW

<https://doi.org/10.34216/1998-0817-2025-31-3-190-194>

ВЫДЕЛЕНИЕ ЯДРА И ПЕРИФЕРИИ ГЛАГОЛОВ ДВИЖЕНИЯ В РУССКОМ ЯЗЫКЕ НА МАТЕРИАЛЕ КОРПУСНЫХ ДАННЫХ

Хохлова Мария Владимировна, кандидат филологических наук, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, m.khokhlova@spbu.ru, <http://orcid.org/0000-0001-9085-0284>

Аннотация. Распределение лексических единиц в текстах характеризуется неравномерностью и асимметричностью ввиду того, что они встречаются с разной частотностью, поэтому традиционные описательные статистики не всегда могут быть показательными при представлении лексики. В работе дается краткий обзор метрик, которые используются для описания высоко- и низкочастотных единиц в разных областях, среди которых особое внимание уделяется ранговому среднему. В качестве материала для исследования была выбрана лексико-семантическая группа глаголов движения, которая хорошо описана в литературе, однако вопрос об упорядочении единиц внутри нее остается открытым. Автором были проанализированы частотности около 900 лексических единиц из словаря «Лексико-семантические группы русских глаголов» по Национальному корпусу русского языка, которые далее были представлены в виде частотного словаря. При помощи рангового среднего были выделены ядерная и периферийные части списка. Ядерная лексика составляет около 12,74 % от общего количества описанных в словаре глаголов, при этом на них приходится около 70 % всех словоупотреблений. Значения высокочастотных глаголов подразумевают отсутствие направленности движения или, напротив, наличие конечного пункта. Результаты исследования могут быть востребованы при преподавании языка, например, при составлении лексических минимумов.

Ключевые слова: глаголы движения, русский язык, ядро, периферия, ранговое распределение, частотный словарь, ранговое среднее.

Для цитирования: Хохлова М.В. Выделение ядра и периферии глаголов движения в русском языке на материале корпусных данных // Вестник Костромского государственного университета. 2025. Т. 31, № 3. С. 190–194. <https://doi.org/10.34216/1998-0817-2025-31-3-190-194>

Благодарности. Работа выполнена при поддержке СПбГУ, шифр проекта 124032900006-1.

Research Article

IDENTIFICATION OF THE CORE AND PERIPHERY OF VERBS OF MOTION IN THE RUSSIAN LANGUAGE BASED ON CORPUS DATA

Maria V. Khokhlova, PhD in Philology, St Petersburg State University, St. Petersburg, Russia, m.khokhlova@spbu.ru, <http://orcid.org/0000-0001-9085-0284>

Abstract. The distribution of lexical units in texts is characterized by skewness and asymmetry due to the fact that they occur with different frequencies, so traditional descriptive statistics may not always be indicative when presenting vocabulary. The paper provides a brief overview of metrics that are used to describe high- and low-frequency units in different areas, with special attention paid to the rank average. The lexical and semantic group of verbs of motion well described in the literature, was chosen as the material for the study; however, the question of ordering units within it remains open. The author analysed the frequencies of about 900 lexical units from the dictionary “Lexical-semantic groups of Russian verbs” according to the National Corpus of the Russian language, which were then presented in the form of a frequency dictionary. Using the rank average, the core and peripheral parts of the list were identified. The core vocabulary makes up about 12.74 % of the total number of verbs described in the dictionary, while they account for about 70 % of all word usages. The meanings of high-frequency verbs imply the absence of direction of movement or, on the contrary, the presence of a final destination point. The results of the study may be in demand in language teaching, for example, in compiling lexical minimums.

Keywords: verbs of motion, Russian language, core, periphery, rank distribution, frequency dictionary, rank average.

For citation: Khokhlova M.V. Identifying the core and periphery of Russian verbs of motion on corpus data. Vestnik of Kostroma State University, 2025, vol. 31, no. 3, pp. 190–194. (In Russ.) <https://doi.org/10.34216/1998-0817-2025-31-3-190-194>

Acknowledgments. The author acknowledges St. Petersburg State University for a research project 124032900006-1.

Введение. Термин «поле» регулярно используется при описании явлений, демонстрирующих некоторые сходства. Применительно к лексике выделяют лексико-семантические поля (более узко – группы), включающие единицы, связанные парадигматическими отношениями (синонимическими или родовидовыми) [Trier], а также тематической связанностью [Porzig]. Вместе с тем весьма важно разделить множество объектов на те, которые отличаются своим частотным характером, воспроизводимостью, и на остальные – которые, хотя и многочисленны, но характеризуются низкой встречаемостью в речи.

Выделение ядра и периферии. Лексика может быть снабжена количественными характеристиками, одной из которых является количество употреблений той или иной единицы в тексте (корпусе), – это дает возможность упорядочить лексические данные в виде частотного словаря. В свою очередь, распределение лексики в частотном словаре отличается асимметричностью и описывается эмпирическим законом Ципфа, связывающим частоту слова и его ранг (цит. по: [Мартыненко 2019]):

$$k_r = \frac{k_{\max}}{r^\gamma},$$

где r – ранг слова, k_r – частота слова ранга r , $k_{\max}$ – частота самого частого слова, γ – коэффициент, характеризующий неравномерность распределения частот. Значения $k_{\max}$ и γ сильно варьируются в зависимости от языков и жанров.

Частота n -го слова в тексте примерно обратно пропорциональна его рангу: оно встречается в n раз реже, чем самое частотное слово (первое в ранжированном списке). В связи с этим замечанием для исследовательских целей важным является выделение высоко- и низкочастотных частей, которые позволили бы отнести лексику к данным категориям.

Поскольку частотный словарь можно представить в виде рангового распределения, это позволяет применять к количественным данным соответствующие статистические метрики. Некоторые из них используются для упорядочения лингвистических данных, а то время как другие применяются на ином материале. Так, индекс Парето [Pareto] связан с распределением доходов среди населения, для оценки которых также был предложен индекс Джини [Джини]. Закон Брэдфорда [Bradford] показывает распределение журналов в зависимости от количества статей с последующим выделением нескольких групп (ядра и других зон), которое будет описываться соотношением $1 : n : n^2$, где $n > 1$. Широко известный в наукометрии индекс Хирша был введен для определения продуктивности ученых согласно количеству цитирований их научных работ [Hirsch]. Представлению языковых данных посвящена работа [Мартыненко, Григорьев],

в которой дан обзор индексов концентрации частотных словарей.

В качестве метрики, позволяющей разделить частотный словарь на ядро и хвост, в работе [Мартыненко 1988] вводится понятие рангового среднего, которое вычисляется по следующей формуле:

$$r = \sum \frac{r^* f_r}{N},$$

где r – ранг лексемы в частотном списке, f_r – соответствующая этому рангу частота, N – объем выборки.

Ранговое среднее можно рассматривать в качестве показателя концентрации частот, которые будут в таком случае расположены в верхней зоне частотного распределения [Мартыненко 2019: 118], что дает возможность использовать полученное значение в качестве границы, разделяющей ранговое распределение на ядро и хвост или на ядро и периферию применительно к лексическим единицам.

Материал исследования. В качестве материала для исследования нами была выбрана группа глаголов движения (перемещения в пространстве), поскольку входящие в нее лексические единицы отличаются конкретными значениями, а также своей многочисленностью, при этом вопрос об их количестве остается открытым.

В Академической грамматике русского языка приводятся 18 глагольных пар, которые отличаются значением однократности – многократности движения, а также его однонаправленности – неоднаправленности [АГ-80]: *бежать – бегать*, *брести – бродить*, *везти – возить*, *вести – водить*, *гнать – гонять*, *гнаться – гоняться*, *ехать – ездить*, *идти – ходить*, *катить – катать*, *катиться – кататься*, *лечь – лезть*, *лететь – летать*, *нести – носить*, *нестись – носиться*, *плыть – плавать*, *ползти – ползать*, *тащить – таскать*, *тащиться – таскаться*.

Глаголы движения также находят отражение в иных источниках: в словарях-справочниках, которые служат как цели структурирования лексических единиц, так и учебным целям. Наиболее полно глаголы движения представлены в идеографическом словаре [Лексико-семантические группы], отражающем семантическую классификацию русской глагольной лексики, которая насчитывает около 2,5 тыс. единиц. На верхнем уровне иерархии оказываются три семантических поля (действия, состояния и отношения), которые далее делятся на подполя. Минимальной единицей описания является отдельная лексико-семантическая группа (ЛСГ), глаголы внутри которой имеют общую архисему, а также проявляют типовую сочетаемость. Всего в словаре описано 108 ЛСГ. При этом глаголы представлены в словаре «в своих основных значениях» [Лексико-семантические группы: 5], которые указаны первыми в толковых слова-

рях. Для каждой ЛСГ авторами выделяются базовые лексемы при помощи процедуры *ступенчатой идентификации*, которая позволяет поэтапно свести значение рассматриваемого глагола к более общему [Лексико-семантические группы]. Для видовой пары указывается только та форма, которая в словаре [МАС] имеет прямое толкование.

В указанном словаре [Лексико-семантические группы] глаголы движения описаны в подполе перемещения, к которой относятся несколько групп¹: 1) ЛСГ глаголов перемещения в пространстве (*передвигать, передвигаться, перемещать, перемещаться*); 2) ЛСГ глаголов вертикального перемещения (*опускать, опускаться, поднимать, подниматься, ронять*); 3) ЛСГ глаголов перемещения, ориентированного относительно конечного пункта (*доставить, достигнуть, приблизиться, придвинуть*); 4) ЛСГ глаголов перемещения, ориентированного относительно исходного пункта (*отдалить, отдалиться, удалить, удалиться*); 5) ЛСГ глаголов перемещения, ориентированного относительно какого-либо промежуточного пункта (*миновать, обвести, обходить, перевести, переходить, провести, проходить*). Часть глаголов описана в нескольких группах: например, *перелететь / перелетать*, принадлежащие первой и пятой из указанных групп. Всего в данном подполе представлено 895 уникальных глаголов, при этом видовые пары приведены вместе (например, *скользить / скользнуть, разъехаться / разъезжаться, уплыть / уплыть*), в то время как упомянутые выше глагольные пары, отличающиеся кратностью или направленностью движения, указаны раздельно (например, *ехать и ездить, идти и ходить* или *нести и носить*). В дальнейшей работе для пар глаголов совершенного и несовершенного вида были подсчитаны частотности для всех вхождений.

Постановка задачи и методология исследования. Приведенный в словаре [Лексико-семантические группы] список является исчерпывающим,

однако нуждается в упорядочении. В нашей работе используется описанная выше метрика рангового среднего, которая была применена для разделения списка лексем на ядерную и периферийную части. Частотности глаголов были получены по данным Национального корпуса русского языка [НКРЯ] в единицах ipm. Нами был рассмотрен подкорпус со снятой вручную грамматической неоднозначностью, при поиске результатов были исключены причастия.

Результаты. Прежде всего, обращает на себя внимание тот факт, что около 4 % лексем имеют нулевое количество вхождений в подкорпусе: *взмываться, выскользнуть, добрасывать, добросить, докидывать, доскакивать, иммигрировать, обваливать, обвезти, обкатить, обскакивать, обсыпаться, обсыпаться, отгрести, отгрести, откатывать, откочевать, отсаживаться, отсесть, переволакивать, перекачивать, перелечь, переплывать, перетянуться, подступаться, подчаливать, подчалить, пригрести, приземлять, приползть, прискакивать, промелькивать, пропалзывать, просыпать, развеивать, стряхать, стряхать, упархивать*. Среди приведенных примеров наибольшее количество единиц приходится на глаголы, предполагающие наличие субъекта, объекта, а также имеющие сему вертикального движения. Все перечисленные глаголы с нулевой частотностью были исключены из дальнейшего рассмотрения.

В таблице 1 приведены значения описательных статистик для глаголов в каждой подгруппе ЛСГ субъектного, объектного и субъектно-объектного перемещения. Максимальные значения данных метрик показаны для подгруппы субъектного перемещения относительно конечного пункта (в целом это наиболее частотные лексические единицы). Однако, как можно заметить, значения отличаются большим разбросом, и, следовательно, упомянутые описательные статистики не являются показательными.

Для каждой приведенной в словаре ЛСГ отдельно были выделены ядерная часть и периферия посред-

Таблица 1

Значения описательных статистик

	Первая группа глаголов (перемещение в пространстве)		Вторая группа глаголов (вертикальное перемещение)		Третья группа глаголов (перемещение относительно конечного пункта)		Четвертая группа глаголов (перемещение относительно исходного пункта)		Пятая группа глаголов (перемещение относительно промежуточного пункта)	
	субъект.	объект. и субъект.	субъект.	объект. и субъект.	субъект.	объект. и субъект.-объект.	субъект.	объект. и субъект.-объект.	субъект.	объект. и субъект.-объект.
среднее (ipm)	45,05	31,16	25,80	28,96	71,38	25,84	42,10	16,45	27,27	35,81
стандартное отклонение (ipm)	131,88	54,21	45,44	51,10	146,30	50,37	121,05	20,21	85,80	69,36

Таблица 2

Количественные показатели для ядерной лексики

	Первая группа глаголов (перемещение в пространстве)		Вторая группа глаголов (вертикальное перемещение)		Третья группа глаголов (перемещение относительно конечного пункта)		Четвертая группа глаголов (перемещение относительно исходного пункта)		Пятая группа глаголов (перемещение относительно промежуточного пункта)	
	субъект.	объект. и субъект.	субъект.	объект. и субъект.	субъект.	объект. и субъект.-объект.	субъект.	объект. и субъект.-объект.	субъект.	объект. и субъект.-объект.
ранговое среднее	10,58	10,21	11,14	6,10	7,80	9,49	6,22	8,29	5,14	2,21
ядро (%)	10,22	15,15	21,62	22,58	12,17	15,13	10,81	23,29	12,66	15,38
частотность ядра (%)	74,74	66,77	66,64	65,37	70,98	67,27	72,17	63,09	71,54	71,31

ством вычисления рангового среднего: чем больше значение данной величины, тем более равномерным является распределение. И, наоборот, меньшие значения свидетельствуют о частотной концентрации, то есть указывают на зону, в которой содержатся лексические единицы, характеризующиеся высокими количественными показателями. Ниже будут рассмотрены ЛСГ, представленные в словаре, а также общий список глаголов.

В таблице 2 приведены значения рангового среднего, доля глаголов, составляющих ядро, и их частотность в каждой подгруппе ЛСГ субъектного, объектного, а также субъектно-объектного перемещения.

Наибольшая доля кумулятивной частоты (74,74 %) показана ядром глаголов первой группы, связанных с перемещением в пространстве. Для данной ЛСГ было выделено 10 ядерных лексем, к которым относятся следующие: *бежать, двигаться / двинуться, ездить, ехать, идти, лететь, следовать, уехать / уезжать, уйти / уходить, ходить*. Они составляют 10,3 % от общего списка глаголов данной ЛСГ с общей частотностью 74,74 % от общего объема примеров, что является максимумом в сопоставлении с единицами иных ЛСГ.

Отметим отсутствие корреляции между ранговым средним и количеством единиц, которые могут быть причислены к ядерным. Ядро подгруппы второй ЛСГ (субъектное перемещение) глаголов вертикального перемещения имеет наиболее высокое значение рангового среднего (11,14) и при этом сформировано наибольшим количеством единиц за счет видовых пар (24 лексемы): *вскочить / вскакивать, лезть, подняться / подниматься (подыматься), спуститься / спускаться* и др. Это позволяет выдвинуть предположение о том, что частотное распределение лексических единиц данной группы характеризуется большей равномерностью и менее выраженным от-

рывом ядра от хвостовой части. В то же время между ядром и частотностью наблюдается сильная отрицательная корреляция (коэффициент корреляции Спирмена равен $-0,9$, $p\text{-value} \ll 0,05$): чем меньше ядро, тем больше количество примеров в корпусе с входящими в него глаголами.

Наибольшую разность в концентрации частот демонстрирует подгруппа пятой ЛСГ (объектное и субъектно-объектное перемещение), подразумевающая у глаголов наличие семы промежуточного пункта и характеризующаяся также своей малочисленностью: ранговое среднее равно 2,21, что дало возможность включить в ядро только две видовые пары (*привести / проводить* и *перевести / переводить*).

Для общего частотного списка глаголов значение ранговой средней составило 63,69, что, таким образом, позволило получить перечень из 114 единиц (включающий в большинстве случаев видовые пары), – ядерная лексика составляет около 12,74 % от общего количества описанных в словаре глаголов. К ним относятся, например, такие лексемы, как *идти, ходить, прийти / приходить, выйти / выходить, пройти / проходить, лететь, вступить / вступать* и др. Средняя частотность входящих в ядро данного распределения глаголов составляет 193,04 ipm, при этом на них приходится 69,38 % всех словоупотреблений. Более трети (34,21 %) лексических единиц содержат сему конечного пункта (*вернуться, достигнуть / достигать, приблизиться / приближаться, принести / приносить, явиться / являться*), в то время как около одной пятой части (22,81 %) – перемещение в пространстве безотносительно к какому-либо месту (*бежать, гулять, ездить, ехать, уехать / уезжать*).

Заключение. Распределение глаголов, принадлежащих ЛСГ движения (перемещения) в пространстве, отличается большим размахом частот, а также

неоднородностью, что затрудняет выделение ядерной и периферийной частей для этой группы лексических единиц. Используемая метрика рангового среднего позволила определить часть распределения, в которых сконцентрированы лексемы, отличающиеся высокими частотностями: их доля составляет более 12 % от всех глаголов, представленных в рассмотренном нами словаре. Значения входящих в него высокочастотных глаголов подразумевают отсутствие направленности движения или, напротив, наличие конечного пункта. Результаты исследования могут быть востребованы при преподавании языка, например при составлении лексических минимумов.

Примечание

¹ Здесь и далее в скобках приведены базовые глаголы согласно [Лексико-семантические группы] для разных групп, предложенная нумерация групп далее будет использована в статье.

Список литературы

Источники

Лексико-семантические группы русских глаголов: учеб. словарь-справ. / под общ. ред. Т.В. Матвеевой. Свердловск: Изд-во Урал. ун-та, 1988. 153 с.

Национальный корпус русского языка. URL: <http://ruscorpora.ru> (дата обращения: 06.06.2025). (НКРЯ)

Русская грамматика. Т. 1: Фонетика. Фонология. Ударение. Интонация. Словообразование. Морфология / Н.Ю. Шведова (гл. ред.). Москва: Наука, 1980. 789 с. (АГ-80)

Словарь русского языка: в 4 т. / АН СССР, Ин-т рус. яз.; под ред. А.П. Евгеньевой. 2-е изд., испр. и доп. Москва: Русский язык, 1981–1984. (МАС)

Исследования

Джини К. Средние величины. Москва: Статистика, 1970. 448 с.

Мартыненко Г.Я. Основы стилистики. Ленинград: Изд-во Ленинград. ун-та, 1988. 173 с.

Мартыненко Г.Я. Методы математической лингвистики в стилистических исследованиях. Санкт-Петербург: Нестор-История, 2019. 296 с.

Мартыненко Г.Я., Григорьев Ю.Д. Индексы концентрации частотных словарей // Вестник Новосибирского государственного университета. Сер.: Лингвистика и межкультурная коммуникация. 2017. Т. 15, № 1. С. 41–53.

Bradford S.C. Sources of Information on Specific Subjects. London, Engineering: An Illustrated Weekly Journal, 1934 (26 January), vol. 137, pp. 85–86.

Hirsch J.E. An index to quantify an individual's scientific research output. PNAS, 2005, no. 102 (46), pp. 16569–16572. URL: <http://www.pnas.org/content/102/46/16569.full> (access date: 06.06.2025).

Pareto W. Cours d'economie politique. Vol. 1–2. Lausanne, 1896–1897.

Porzig W. Wesenhafte Bedeutungsbeziehungen. Beiträge zur Geschichte der Deutschen Sprache und Literatur, 1934, no. 58, pp. 70–97.

Trier J. Der Deutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines Sprachlichen Feldes. Heidelberg, Winter, 1931, Bd. 1.

References

Dzhini K. *Srednie velichiny* [The averages]. Moscow, Statistika Publ., 1970, 448 p. (In Russ.)

Martynenko G.Ia. *Osnovy stilemetrii* [Foundations of stylometry]. Leningrad, Izd-vo Leningrad. un-ta Publ., 1988, 173 p. (In Russ.)

Martynenko G.Ia. *Metody matematicheskoi lingvistiki v stilisticheskikh issledovaniakh* [Methods of mathematical linguistics applied to research in stylistics]. St. Petersburg, Nestor-Istoriia Publ., 2019, 296 p. (In Russ.)

Martynenko G.Ia., Grigor'ev Yu.D. *Indeksy kontsentratsii chastotnykh slovarei* [Concentration indices of frequency dictionaries]. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Ser.: Lingvistika i mezhkul'turnaia kommunikatsiia* [Bulletin of the Novosibirsk State University. Ser.: Linguistics and intercultural communication], 2017, vol. 15, no. 1, pp. 41–53. (In Russ.)

Bradford S.C. Sources of Information on Specific Subjects. London, Engineering: An Illustrated Weekly Journal, 1934 (26 January), vol. 137, pp. 85–86.

Hirsch J.E. An index to quantify an individual's scientific research output. PNAS, 2005, no. 102 (46), pp. 16569–16572. URL: <http://www.pnas.org/content/102/46/16569.full> (access date: 06.06.2025).

Pareto W. Cours d'economie politique. Vol. 1–2. Lausanne, 1896–1897.

Porzig W. Wesenhafte Bedeutungsbeziehungen. Beiträge zur Geschichte der Deutschen Sprache und Literatur, 1934, no. 58, pp. 70–97.

Trier J. Der Deutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines Sprachlichen Feldes. Heidelberg, Winter, 1931, Bd. 1.

Статья поступила в редакцию 06.06.2025; одобрена после рецензирования 15.07.2025; принята к публикации 16.07.2025.

The article was submitted 06.06.2025; approved after reviewing 15.07.2025; accepted for publication 16.07.2025.