

**ВЕРИФИКАЦИЯ ПРОЦЕДУРЫ  
ПОЛУЧЕНИЯ ПАРАМЕТРИЧЕСКОГО ЯДРА  
(НА МАТЕРИАЛЕ СЛОВАРЯ РУССКОГО ЯЗЫКА В 4 ТОМАХ  
ПОД РЕДАКЦИЕЙ А. П. ЕВГЕНЬЕВОЙ)**

**Ю. А. Стародубцева**

*Воронежский государственный университет*

Поступила в редакцию 24 декабря 2016 г.

**Аннотация:** *статья посвящена анализу метода слияния ядер и метода слияния ненулевой лексики. В результате исследования оценивается эффективность различных методов выделения ядра лексики.*

**Ключевые слова:** *параметрический анализ лексики, лексико-семантическое ядро, словарь, метод, параметрический вес, доминанта.*

**Abstract:** *the article compares two methods of core vocabulary analysis: the method of merging of cores and the method of the non-zero vocabulary merging. As a result the efficiency of different methods of allocating the core vocabulary of the Russian language is evaluated.*

**Key words:** *parametric analysis of vocabulary, lexico-semantic core, dictionary, method, parametric weight, dominant lexeme.*

Метод параметрического анализе лексики (далее – ПАЛ) был предложен и апробирован на материале романских языков В. Т. Титовым. Он подробно описан в работах [1–5], и здесь нет необходимости воспроизводить его описание.

В. Т. Титов рассматривал два метода получения параметрического ядра (далее – ПЯ) лексики: «При решении вопроса о технике суммирования весов возникает потребность рассмотреть следующие теоретические проблемы. <...> ...слияние четырех ядер размеров в 1000 (и менее) слов – значительно менее трудоемкое дело, чем слияние четырех выборок по 10 000–12 000 слов. Правда, при слиянии ядер некоторые лексемы могут терять в своем суммарном параметрическом весе, если по какому-то из параметров они не попали в ядро. Это может привести к искажению реального суммарного параметрического веса отдельных лексем.

<...> ...на французском материале было проведено специальное исследование по определению суммарного параметрического веса двумя различными способами: *слиянием четырех ядерных списков и слиянием четырех полных версий словаря* (выделено нами. – Ю. С.) Эксперимент показал, что расхождения в точности полученных результатов ни в коей мере не сопоставимы с расхождением в трудоемкости двух способов: малый (если не сказать *незначительный*) выигрыш в точности не оправдывает тех колоссальных затрат труда и времени, которыми он приобретает» [5, с. 185].

«Слияние четырех ядерных списков» назовем *Методом слияния ядер* (далее – МСЯ), а «слияние четырех полных версий словаря» близко к предлагаемому нами *Методу слияния ненулевой лексики* (далее – МСНЛ). Отличие состоит в том, что мы не видим необходимости включать во множества, подлежащие слиянию, 1) самые длинные слова (если они не имеют положительных весов по другим параметрам, кроме функционального), 2) однозначные слова, 3) слова, не имеющие в словаре фразеосочетаний, и 4) слова, не имеющие синонимов.

Эксперимент на материале франко-русского словаря, упоминающийся В. Т. Титовым, к сожалению, на нашел отражения в его диссертации и не был опубликован впоследствии. До настоящего времени в ПАЛ для получения ПЯ лексики использовался исключительно метод слияния ядер.

По существу, отношение к Методу слияния ненулевых ядер (далее – МСНЯ) на сегодняшний день остается вопросом веры: верить или не верить утверждению В. Т. Титова. Однако наука, в отличие от религии, основана не на вере, а на сомнении, и вопрос об оценке МСНЯ в науке остается открытым.

Цель нашей статьи и состоит в сопоставлении МСЯ и МСНЛ и верификации их эффективности на фоне других опытов выделения ядра русской лексики, как в рамках параметрического анализа лексики, так и вне его.

Объектами сравнения являются лексико-семантические ядра, представляющие единства «Словарь + Метод».

Для оценки эффективности предложенного метода слияния ненулевой лексики было проведено сравнение 7 различных источников:

1) Русский ассоциативный словарь (6012 слов) [6]; Частотный словарь + ассоциативный метод;

2) Частотный словарь современного русского языка О. Н. Ляшевской и С. А. Шарова (995 слов) [7]; (НКРЯ + квантитативный метод);

3) Лексическая основа русского языка В. В. Морковкина (2281 слово) [8]; (Частотные словари + педагогический метод);

4) параметрическое ядро по данным словаря С. И. Ожегова<sup>1</sup> (1001 слово) [9]; (Словарь Ожегова + МСЯ);

5) сводное (по нескольким словарям)<sup>2</sup> параметрическое ядро русского языка (1000 слов) [4]; (русско-финский словарь – Ф-параметр, словаря Ушакова – Д-параметр, электронный фразеологический и синонимический словаря русского языка – С- и П-параметры, соответственно, + МСЯ).

6) параметрическое ядро русского языка, полученное методом слияния ядер по данным Словаря русского языка в 4 томах под редакцией А. П. Евгеньевой (далее – МАС-2) (1001 слово) [10]; (МАС-2 + МСЯ).

7) параметрическое ядро русского языка, полученное методом слияния ненулевой лексики (1001 слово); (МАС-2 + МСНЛ).

Первое параметрическое ядро русского языка по данным МАС-2 получено нами посредством МСЯ. Анализ лексики МАС-2 по четырем системообразующим параметрам позволил получить четыре ядра русской лексики: функциональное ядро, насчитывающее 1145 слов, синтагматическое ядро – 1091 слово, парадигматическое ядро – 1529 слов, эпидигматическое ядро – 1085 слов.

В результате слияния четырех ядер русской лексики и сложения четырех частнопараметрических весов представленных в них слов мы получили пара-

<sup>1</sup> Благодарим А. А. Кретова за предоставленные данные по ПЯ словаря С. И. Ожегова.

<sup>2</sup> «...Малые инославяно-русские словари ...оптимальны лишь по одному из параметров – функциональному. В остальных случаях <...> лучшие результаты дает обращение к словарям синонимов (парадигматика), фразеологии (синтагматика) и одноязычным толковым словарям (эпидигматика – многозначность). Привлечение соответствующих источников помогает корректировать результаты. Для русского и словацкого языков пришлось получить также по два ядра. Для русского: первый раз – с использованием данных словаря Ожегова (№ 15), второй раз – с использованием данных русско-финского словаря (Ф-параметр), словаря Ушакова (№ 19: Д-параметр), а также электронных фразеологического и синонимического словарей русского языка (С- и П-параметры соответственно)» [4, с. 53–54]. Благодарим И. А. Меркулову за предоставленные данные по русскому сводному параметрическому языку.

метрические веса лемм. При этом все слова, вошедшие хотя бы в одно ядро, помимо функционального, получили функциональный вес. Это отличает наш подход от классического, в связи с чем назовем его не просто *Методом слияния ядер* (МСЯ), а *Методом слияния ядер – модернизированным* (МСЯм).

Сортировка лемм в порядке убывания (суммарно) параметрического веса дала следующие результаты.

По данным ПЯ МАС-2 глагол *бить* имеет максимальный параметрический вес и является доминантой русской лексики. На одну стотысячную от него отстает вице-доминанта – глагол *дать*. Фактически по полученным данным МАС-2 можно говорить о двух доминантах: *бить* и *дать*.

«Верхушка» ПЯ русской лексики представлена в табл. 1.

Согласно МСНЯ для выявления параметрического ядра русской лексики необходимо взять все ненулевые веса лемм по каждому из параметров, затем слить эти четыре множества слов. Соблюдая данный принцип выделения ядра, нет необходимости ориентироваться на количество параметров. Так, в новое параметрическое ядро по синтагматическому параметру попадают слова с одним фразеосочетанием и больше, по эпидигматическому – с двумя значениями и больше, по парадигматическому – с двумя синонимами и больше. Функциональный параметр приписывается всем словам, которые входят в ядро.

В результате в новое параметрическое ядро русского языка по данным МАС-2 по парадигматическому параметру вошло 12 242 слова, по синтагматическому – 5291 слово, по эпидигматическому – 26 231 слово, по функциональному – все слова, получившиеся в результате слияния этих трех множеств, – 19 395 слов.

Для внесения дискретности в ядро, полученное методом слияния ненулевой лексики, параметрический вес для каждой леммы округляется до целого числа. В результате получается четыре множества: 1) лексика, имеющая параметрический вес 4 (малое параметрическое ядро); 2) лексика, имеющая параметрический вес 3 (малая периферия); 3) лексика, имеющая параметрический вес 2 (большая периферия); 4) квазирелевантная лексика с параметрическим весом 1.

Так, малое ядро насчитывает 639 слов. Малая периферия представлена 2379 словами. Большая периферия включает 9761. И 19 396 слов – это квазирелевантная лексика с параметрическим весом до 1 (рис. 1).

По данным параметрического ядра МАС-2, полученного методом слияния ненулевой лексики, доминантой является глагол *дать*, а вице-доминантой глагол *бить*. «Верхушка» нового параметрического ядра русской лексики представлена в табл. 2.

Малое параметрическое ядро русского языка по данным МАС-2

Лемма	Дефиниция	Д	П	С	Ф	Σ
БИТЬ	Ударять, колотить	0,999	1,000	0,999	0,987	3,985
ДАТЬ	Передать из рук в руки, вручить	1,000	0,998	1,000	0,987	3,985
ДОМ	Здание, строение, предназначенное для жилья, для размещения различных учреждений и предприятий	0,985	0,998	0,998	0,987	3,968
МИР <sup>1</sup>	Совокупность всех форм материи в земном и космическом пространстве; Вселенная	0,999	0,973	0,998	0,987	3,957
ИДТИ	Передвигаться, перемещаться в пространстве: а) передвигаться, ступая ногами, делая шаги (о человеке и животном)	1,000	0,993	1,000	0,964	3,957
СЕСТЬ	Принять сидячее положение, занять место, предназначенное для сидения	0,999	0,993	0,998	0,964	3,954
ЛИЦО	Передняя часть головы человека	0,989	0,998	1,000	0,964	3,951
БОЙ	Битва, сражение; действие по глаг. «биться» (в 1 знач.)	0,994	0,993	0,975	0,987	3,950
УБИТЬ <sup>1</sup>	Лишить жизни, умертвить	0,992	1,000	0,994	0,964	3,950
РЯД	Совокупность предметов, лиц, расположенных один к одному, друг за другом, в одну линию	0,999	0,973	0,987	0,987	3,946
ВЕЩЬ	Всякий отдельный предмет (преимущественно бытового обихода, трудовой деятельности и т.п.)	0,994	0,973	0,987	0,987	3,942
ШУМ	Совокупность неясных глухих звуков, сливающихся в однообразное звучание; гул	0,985	0,993	0,975	0,987	3,941
ВЕРХ	Самая высокая часть чего-л., верхняя часть, оконечность чего-л.; противоп. низ	0,996	0,993	0,987	0,964	3,939
ДЕЛО	Работа, занятие, деятельность	1,000	0,973	1,000	0,964	3,937
ОТЕЦ	Мужчина по отношению к своим детям	0,992	0,993	0,987	0,964	3,936
ЖИЗНЬ	Особая форма движения материи, возникающая на определенном этапе ее развития	0,997	0,973	1,000	0,964	3,933
КРОВЬ	Жидкая ткань, которая движется по кровеносным сосудам организма	0,989	0,973	1,000	0,964	3,926
ВОЛЯ	Одно из свойств человеческой психики, выражающееся в способности добиваться осуществления поставленных перед собой целей	0,989	0,973	0,997	0,964	3,924
СЛЕД <sup>1</sup>	Отпечаток, оттиск ноги или лапы на какой-л. поверхности	0,989	0,973	0,995	0,964	3,922
КРАЙ <sup>1</sup>	Предельная линия, ограничивающая поверхность или протяженность чего-л.	0,985	0,973	0,999	0,964	3,921

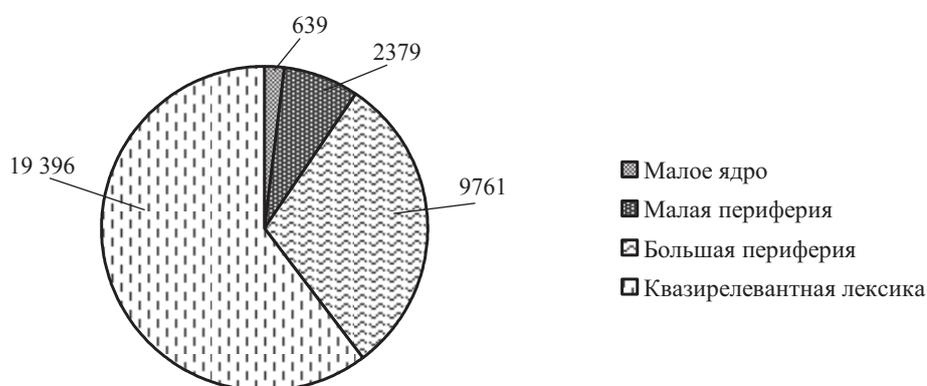


Рис. 1. Параметрическое ядро русской лексики по данным МАС-2, полученное методом слияния ненулевой лексики

Таблица 2

Фрагмент малого параметрического ядра ненулевой лексики русского языка по данным МАС-2

Лемма	Дефиниция	Д	П	С	Ф	Σ	Кол-во Ядер
ДАТЬ	Передать из рук в руки, вручить	0,99963	0,99961	0,99989	0,9873	3,98642	4
БИТЬ	Ударять, колотить	0,99892	0,99576	0,99915	0,9873	3,98111	4
ЖИТЬ	Существовать, быть живым	0,99568	0,99407	0,99531	0,9873	3,97234	4
ПУТЬ	Полоса земли, служащая для езды и ходьбы; дорога	0,99669	0,98738	0,99964	0,9873	3,97099	4
ЛЕЧЬ	Принять горизонтальное положение (о людях и некоторых животных)	0,99850	0,99807	0,98693	0,9873	3,97078	4
БЫТЬ	Существовать	0,98940	0,99407	0,99888	0,9873	3,96963	4
ДУТЬ	Нести, гнать струи воздуха, приводить воздух в движение (о ветре)	0,99236	0,99576	0,99394	0,9873	3,96933	4
ДУХ	Психические способности, сознание, мышление	0,99738	0,98068	0,99929	0,9873	3,96463	4
ИДТИ	Передвигаться, перемещаться в пространстве: а) передвигаться, ступая ногами, делая шаги (о человеке и животном)	0,99995	0,99761	0,99968	0,9635	3,96070	4
УЙТИ	Покинуть какое-л. место, чье-л. общество; удалиться, отправиться куда-л.	0,99976	0,99807	0,99745	0,9635	3,95876	4
СТАТЬ <sup>1</sup>	Принять стоячее положение, подняться на ноги; встать	0,99876	0,99576	0,99947	0,9635	3,95746	4
БОЙ	Битва, сражение; действие по глаг. «биться» (в 1 знач.)	0,99433	0,99761	0,97535	0,9873	3,95456	4
ДОМ	Здание, строение, предназначенное для жилья, для размещения различных учреждений и предприятий	0,98492	0,98068	0,99782	0,9873	3,95070	4
БРАТЬ	Принимать в руки, схватывать руками (зубами, щипцами и т. п.)	0,99980	0,98738	0,99971	0,9635	3,95036	4
ЛИЦО	Передняя часть головы человека	0,98940	0,99407	0,99971	0,9635	3,94665	4
ДЕЛО	Работа, занятие, деятельность	0,99963	0,98068	0,99993	0,9635	3,94371	4
ИМЯ	Личное название человека, даваемое ему при рождении	0,97765	0,98068	0,99701	0,9873	3,94262	4
ЖИЗНЬ	Особая форма движения материи, возникающая на определенном этапе ее развития	0,99669	0,98068	0,99962	0,9635	3,94045	4
КРЫТЬ	Делать над чем-л. верх, покрытие, крышу	0,99236	0,99135	0,98693	0,9635	3,93410	4
КРОВЬ	Жидкая ткань, которая движется по кровеносным сосудам организма и обеспечивает питание его клеток и обмен веществ в нем	0,98940	0,98068	0,99968	0,9635	3,93323	4

Таблица 3

Распределение совпавшей лексики в источниках

Кол-во словарей	Кол-во слов	Накопление	%
7	89	89	1
6	148	237	3
5	212	449	6
4	288	737	9
3	652	1389	18
2	1131	2520	32
1	5350	7870	100

Из табл. 3 видно, что доля совпавшей лексики во всех семи случаях составляет лишь 1 % (89 слов). 3 % совпадения лексики в 6 случаях и т.д. Конкретные примеры совпавшей лексики во всех семи источниках представлены в табл. 4.

Наименьший процент неподтвержденной лексики оказался в частотном словаре Ляшевской–Шарова. Наибольшее количество неподтвержденной лексики – в «Русском ассоциативном словаре». Большой процент (26 %) неподтвержденной лексики оказался и в Ядре-МСНЛ, полученном методом слияния ненулевой лексики, и «Лексической основе русского языка» Морковкина (табл. 5).

Примеры совпавшей лексики в семи источниках\*

Слово	Значение	ЛОРЯ	МСНЛ	МСЯ	Оже-гов	РАС	Рус-вПЯ	Част_ЛШ
БИТЬ	Ударять, колотить	+	+	+	+	+	+	+
БОЙ	Битва, сражение; действие по глаг. «биться» (в 1 знач.)	+	+	+	+	+	+	+
БЫТЬ	Существовать	+	+	+	+	+	+	+
ВЕРНЫЙ	Заслуживающий полного доверия; преданный	+	+	+	+	+	+	+
ВОЙТИ	Идя, двигаясь, проникнуть куда-л., в пределы чего-л	+	+	+	+	+	+	+
ВОЛЯ	Одно из свойств человеческой психики, выражающееся в способности добиваться осуществления поставленных перед собой целей, осуществления стремлений	+	+	+	+	+	+	+
ВЫСОТА	Протяженность по вертикали снизу вверх; высота	+	+	+	+	+	+	+
ГЛАВА	То же, что <i>голова</i> (в 1 знач.)	+	+	+	+	+	+	+
ГЛАВНЫЙ	Самый важный, существенный среди других; основной	+	+	+	+	+	+	+
ГЛАЗ	Орган зрения	+	+	+	+	+	+	+
ГОВОРИТЬ	Пользоваться, владеть устной речью	+	+	+	+	+	+	+
ГОД	Единица летосчисления, промежуток времени, в течение которого Земля совершает один оборот вокруг Солнца; содержит 12 месяцев, отчисляемых с 1-го января	+	+	+	+	+	+	+
ГОРЯЧИЙ	Имеющий высокую температуру; сильно нагретый	+	+	+	+	+	+	+
ДАТЬ	Передать из рук в руки, вручить	+	+	+	+	+	+	+
ДЕД	Отец отца или матери	+	+	+	+	+	+	+
ДЕЙСТВИЕ	Деятельность	+	+	+	+	+	+	+
ДЕЛО	Работа, занятие, деятельность	+	+	+	+	+	+	+
ДЕРЖАТЬ	Взяв в руки (в рот, в зубы и т. п.), не давать выпасть	+	+	+	+	+	+	+
ДОБРЫЙ	Расположенный к людям, отзывчивый, исполненный доброты, сочувствия к ним, готовности помочь	+	+	+	+	+	+	+
ДОМ	Здание, строение, предназначенное для жилья, для размещения различных учреждений и предприятий	+	+	+	+	+	+	+
ДОРОГА	Полоса земли, служащая для езды и ходьбы	+	+	+	+	+	+	+
ДОРОГОЙ	Стоящий больших денег; <i>противоп.</i> дешевый	+	+	+	+	+	+	+
ДУМАТЬ	Размышлять, предаваться раздумью	+	+	+	+	+	+	+
ДУХ	Психические способности, сознание, мышление	+	+	+	+	+	+	+

\* ЛОРЯ – лексическая основа русского языка, РАС – Русский ассоциативный словарь, РусвПЯ – русское сводное параметрическое ядро, Част\_ЛШ – Частотный словарь О. Н. Ляшевской и С. А. Шарова.

Согласно данным табл. 6 наиболее удачные показатели имеет ассоциативный словарь русского языка. Однако необходимо учесть тот факт, что изначально словники, взятые для сравнения, включали разное количество лексики. И «Русский ассоциативный словарь» состоит из наибольшего количества слов. Взяв за 100 % наименьшее ядро, мы получаем совершенно иную картину (табл. 7).

С точки зрения корреляции параметрических ядер, полученных разными методами и по разным словарям-источникам русского языка, традиционной метод слияния лексики – МСЯ – оказывается наиболее достоверным. Так, рис. 2 свидетельствует, что наиболее информативно ядро русского языка, полученное по данным МАС-2. За ним следуют Сводное параметрическое ядро русского языка, Лек-

Т а б л и ц а 5

*Доля неподтвержденной лексики*

Словник	Размер	К-вес Брутто	ДК-вес	1	Доля неподтв. лексики, %
Част_ЛШ	995	3773	9518	54	5
МСЯ	1001	3958	9333	105	10
РусВПЯ	1000	3645	9646	166	17
Ожегов	1001	3609	9682	176	18
ЛОРЯ	2281	6422	13291	583	26
МСНЛ	1001	3247	10044	262	26
РАС	6012	10685	2606	4004	67

Т а б л и ц а 6

*Совпадение ядер в процентном соотношении*

Словник		ЛОРЯ	МСНЛ	МСЯ	Ожегов	РАС	РусВПЯ	Част_ЛШ
	Кол-во слов	2281	1001	1001	1001	6012	1000	995
ЛОРЯ	2281		41 %	54 %	45 %	63 %	42 %	87 %
МСНЛ	1001	41 %		42 %	27 %	54 %	37 %	24 %
МСЯ	1001	54 %	42 %		48 %	69 %	49 %	33 %
Ожегов	1001	45 %	27 %	48 %		63 %	49 %	28 %
РАС	6012	63 %	54 %	69 %	63 %		59 %	78 %
РусВПЯ	1000	42 %	37 %	49 %	49 %	59 %		28 %
Част_ЛШ	995	87 %	24 %	33 %	28 %	78 %	28 %	

Т а б л и ц а 7

Словник		ЛОРЯ	МСНЛ	МСЯ	Ожегов	РАС	РусВПЯ	Част_ЛШ
	Кол-во слов	2281	1001	1001	1001	6012	1000	995
ЛОРЯ	2281		25 %	33 %	28 %	35 %	26 %	53 %
МСНЛ	1001	25 %		42 %	27 %	15 %	37 %	24 %
МСЯ	1001	33 %	42 %		48 %	20 %	49 %	33 %
Ожегов	1001	28 %	27 %	48 %		18 %	49 %	28 %
РАС	6012	35 %	15 %	20 %	18 %		17 %	22 %
РусВПЯ	1000	26 %	37 %	49 %	49 %	17 %		28 %
Част_ЛШ	995	53 %	24 %	33 %	28 %	22 %	28 %	

сическая основа русского языка Морковкина и параметрическое ядро, полученное по словарю Ожегова. Параметрическое ядро русского языка, полученное методом слияния ненулевой лексики, а также «Русский ассоциативный словарь», принимаемый некоторыми за лексическое ядро русского языка, оказываются информативными в наименьшей степени.

Полученные данные также позволяют сделать следующие выводы: 1) метод слияния ядер является более информативным, чем метод слияния ненулевой лексики; 2) метод влияет на результат сильнее, чем различие источников.

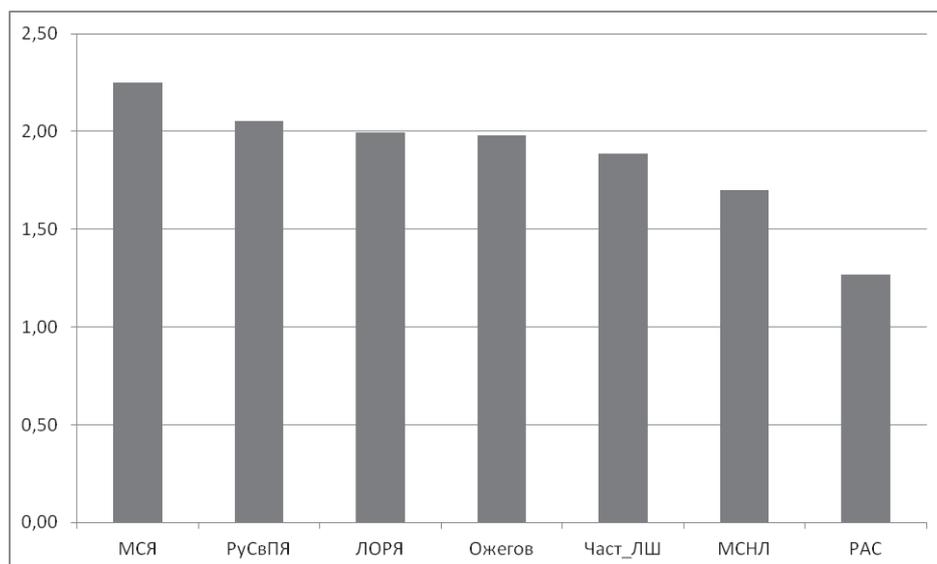


Рис. 2. Корреляция ядерной лексики, рассмотренных в работе источников

#### ЛИТЕРАТУРА

1. *Титов В. Т.* Общая квантитативная лексикология романских языков / В. Т. Титов. – Воронеж : Издательство Воронежского государственного университета, 2002. – 240 с.
2. *Титов В. Т.* Частная квантитативная лексикология романских языков / В. Т. Титов. – Воронеж : Издательство Воронежского государственного университета, 2004. – 552 с.
3. *Титов В. Т.* Квантитативная лексикология романских языков : дис. ... д-ра филол. наук / В. Т. Титов. – Воронеж : Издательство Воронежского государственного университета, 2004б. – 325 с. + Приложения.
4. *Кретов А. А.* Проблемы квантитативной лексикологии славянских языков / А. А. Кретов, И. А. Меркулова, В. Т. Титов // Вопросы языкознания. – 2011. – № 1. – С. 52–65.
5. *Титов В. Т.* Принципы квантитативной лексикологии (на примере романских языков) : дис. ... д-ра филол. наук / В. Т. Титов. – Воронеж, 2005. – 310 с.

6. Русский ассоциативный словарь : в 2 т. Т. 1. От стимула к реакции : ок. 7000 стимулов / Ю. Н. Караулов [и др.]. – М. : Астрель : АСТ, 2002. – 784 с.
7. *Ляшевская О. Н., Шаров С. А.* Частотный словарь современного русского языка (на материалах Национального корпуса русского языка) / О. Н. Ляшевская, С. А. Шаров. – М. : Азбуковник, 2009.
8. Лексическая основа русского языка : комплексный учебный словарь / под ред. В. В. Морковкина. – М. : Русский язык, 1984. – 1168 с.
9. *Ожегов С. И.* Словарь русского языка / С. И. Ожегов ; под ред. Н. Ю. Шведовой. – 23-е изд., испр. – М. : Русский язык, 1991. – 916 с.
10. Словарь русского языка : в 4 т. / под ред. А. П. Евгеньевой; Российская академия наук, Институт лингвистических исследований. – 4-е изд., стер. – М. : Русский язык ; Полиграфресурсы, 1999.

*Воронежский государственный университет  
Стародубцева Ю. А., преподаватель кафедры теоретической и прикладной лингвистики  
E-mail: SuvorovaJ84@mail.ru  
Тел.: 8-903-655-09-93*

*Voronezh State University  
Starodubtseva Yu. A., Lecturer of the Theoretical and Applied Linguistics Department  
E-mail: SuvorovaJ84@mail.ru  
Tel.: 8-903-655-09-93*