

## ФИЛОСОФИЯ | PHILOSOPHY

УДК 130.2

doi:10.35853/vestnik.gu.2025.13-3.10

5.7.8.

### Цифровые инструменты анализа языковых средств в исследовании корпуса текстов Русской идеи

Анастасия Валентиновна Бабаева<sup>1</sup>, Артем Маратович Хамидулин<sup>2</sup>,  
Ольга Валентиновна Смышляева<sup>3</sup>

<sup>1,2,3</sup>Нижегородский государственный педагогический университет  
им. Козьмы Минина, Нижний Новгород, Россия

<sup>1</sup>dff1890@yandex.ru, <https://orcid.org/0000-0002-0676-6336>

<sup>2</sup>am-khamidulin@yandex.ru, <https://orcid.org/0000-0001-7741-9940>

<sup>3</sup>smyshlologa@yandex.ru, <https://orcid.org/0000-0002-0543-1825>

**Аннотация.** Статья посвящена применению цифровых инструментов корпусного анализа языковых средств в процессе изучения текстов Русской идеи. На основе сопоставления существующих программ корпусного анализа текстов авторами статьи была выбрана программа Voyant Tools, с помощью которой выявлены количественные характеристики текстов корпуса Русской идеи, в частности: частота слов и словосочетаний в 120 авторских произведениях. Полученные топ-50 самых частотных слов корпуса Русской идеи были сопоставлены с топ-50 частотных слов из Национального корпуса русского языка. В результате были выявлены 22 совпадения, причем 16 слов являются именами существительными, 4 – именами прилагательными, 2 – глаголами. Интерпретация полученных результатов базировалась на гипотезе, что смысловое ядро Русской идеи должно пересекаться с лексико-семантическим полем русской культуры, зафиксированным в Национальном корпусе русского языка. Семантико-грамматический анализ совпадающих слов позволил не только схематично реконструировать национальную картину мира, но и сформулировать базовый когнитивный сценарий отечественной культуры. Основная интенция данного сценария – нацеленность на созерцание в противовес активистско-деятельностному отношению к миру. Полученные результаты в целом подтверждают имеющиеся исследования по тематике Русской идеи спекулятивного характера и вполне могут использоваться в качестве инструмента фундирования последних.

**Ключевые слова:** Русская идея, история русской философии, Национальный корпус русского языка, цифровые инструменты, корпусный анализ, когнитивный сценарий

**Для цитирования:** Бабаева А. В., Хамидулин А. М., Смышляева О. В. Цифровые инструменты анализа языковых средств в исследовании корпуса текстов Русской идеи // Вестник Гуманитарного университета. 2025. Т. 13, № 3. С. 154–165. DOI 10.35853/vestnik.gu.2025.13-3.10.

## Digital Tools for Analyzing Language Tools in the Study of the Corpus of Texts Related to the Russian Idea

Anastasia V. Babaeva<sup>1</sup>, Artem M. Khamidulin<sup>2</sup>,

Olga V. Smyshliaeva<sup>3</sup>

<sup>1,2,3</sup>Minin Nizhny Novgorod State Pedagogical University (Minin University),

Nizhny Novgorod, Russia

<sup>1</sup>dff1890@yandex.ru, <https://orcid.org/0000-0002-0676-6336>

<sup>2</sup>am-khamidulin@yandex.ru, <https://orcid.org/0000-0001-7741-9940>

<sup>3</sup>smyshlologa@yandex.ru, <https://orcid.org/0000-0002-0543-1825>

**Abstract.** The article is devoted to the utilization of digital instruments in the context of corpus-based analysis of linguistic elements in the examination of texts related to the Russian idea. Based on a comparison of existing corpus analysis programs, the authors of the article selected the Voyant Tools program, which was used to identify the quantitative characteristics of the texts of the corpus of the Russian idea, in particular: the frequency of words and phrases in 120 original works. The top 50 most frequently used words from the Russian Idea corpus (CRI) were compared to the top 50 frequently used words from the National Corpus of the Russian Language (NCRL). As a result, 22 matches were identified, of which 16 words are nouns, 4 are adjectives, and 2 are verbs. The interpretation of the results was based on the assumption that the semantic core of the Russian concept should intersect with the lexical and semantic field of Russian culture as recorded in the National Corpus of the Russian Language. The semantic-grammatical examination of corresponding words not only facilitated the creation of a schematic representation of the national worldview but also enabled the formulation of a fundamental cognitive framework for domestic culture. The primary intention of this scenario is to emphasize introspection rather than an active and action-oriented approach to engagement with the world. The findings obtained generally support the existing research on the topic of the theoretical nature of the Russian concept and can be effectively employed as a means of supporting it.

**Keywords:** Russian idea, history of Russian philosophy, National corpus of the Russian language, digital tools, corpus analysis, cognitive scenario

**For citation:** Babaeva AV, Khamidulin AM, Smyshliaeva OV. Digital Tools for Analyzing Language Tools in the Study of the Corpus of Texts Related to the Russian Idea. *Vestnik Gumanitarnogo universiteta = Bulletin of Liberal Arts University*. 2025;13(3):154-165. (In Russ.). DOI:10.35853/vestnik.gu.2025.13-3.10.

### Введение

Изучение текстов, посвященных Русской идее, обычно базируется на компаративистском и герменевтическом подходах. Первый предполагает сопоставление взглядов различных авторов в конкретных историко-культурных контекстах с целью выстраивания общего национального дискурса в свете этапов его развития. В реконструированном метанарративе выделяются ключевые философемы и стратагемы, между которыми впоследствии с помощью герменевтики воссоздаются взаимосвязи и простраивается иерархия смыслов. При этом часто используется принцип предположения, что предполагает формулировку исходного тезиса, определяющего всю последующую цепочку смыслов. В противовес традиционным исследовательским стратегиям мы предлагаем подойти к анализу Русской идеи с помощью корпусного анализа языковых средств. В этой связи мы намеренно стремимся воздерживаться от изложения пролегоменов Русской идеи и как бы «выносим за скобки», насколько это возможно, до получения эмпирических результатов ее содержательный аспект. Такой подход позволяет представить метанарратив со структурных позиций, где единицами анализа оказываются слова, а не теоретические положения.

В рамках данного исследования мы обращаемся к языку с позиции формалистского подхода, согласно которому язык следует рассматривать в качестве ментального феномена, одновременно являющегося и инструментом познания, и символическим кон-

тейнером хранения его результатов. При этом языковая система представляет собой семантическое поле возможностей построения высказываний и саму совокупность высказываний. Относительно Русской идеи, которую мы, вслед за В. М. Межуевым, содержательно понимаем «как выражение... культурной преемственности в духовной истории народа» [Межуев 1997, с. 8–9], в данном исследовании предлагаем с позиции формы рассматривать в качестве определенного корпуса высказываний, представляющих метанарратив отечественной культуры, которая посредством языковых возможностей манифестирует смыслы и стратегии бытия русского человека как представителя определенного культурного пространства.

Исследователь русской идеи В. Н. Сагатовский называет три источника формирования национального самосознания: народ, интеллигенцию, представителей государственной власти [Сагатовский 1994]. Среди источников, которые мы используем для анализа в данной статье, 120 произведений, авторы которых более чем на 87 % относятся именно к интеллигенции (представителям интеллектуальных профессий, чаще из академической среды), остальной объем текстов распределяется среди представителей народа и государственной власти.

Поиск текстов для исследования происходил по принципу репрезентативности. Сначала были отобраны все работы, где авторы размышляли на тему русского самосознания, затем была произведена сортировка наиболее существенных для нашей темы текстов, количество которых приближается к исчерпывающему с точки зрения статистической значимости и целей данной работы, что подтверждается предыдущим исследованием [Хамидулин, Бабаева 2024]. В результате был сформирован корпус произведений, включающий в себя тексты, созданные в период с XI по XXI в. Полученный таким образом набор произведений рассматривается без погружения в содержание затрагиваемых вопросов и динамических изменений проблематики, а также без учета исторического контекста, в котором создавались сами тексты. История становления Русской идеи, условия ее развития, концептуальные основания данного метанарратива – все эти вопросы уже рассмотрены другими авторами [Криницкая 2013; Кочеров, Парилов 2022; Орлов 2011; Сыромятников 2014], поэтому не затрагиваются в представленном исследовании.

Целью данного исследования является попытка реконструировать базовый когнитивный сценарий русской культуры, зафиксированный в текстах, определяющих корпус Русской идеи. Методы сбора данных для исследования будут описаны ниже, но именно с их помощью мы сможем обозначить топ самых частотных слов для определения реперных точек русской культуры – системы значений, иначе – базового когнитивного сценария. Для реализации цели исследования нам необходимо определить топ самых частотных слов корпуса Русской идеи (КРИ) и сопоставить полученные данные с Национальным корпусом русского языка (НКРЯ). Ведущей гипотезой исследования мы закладываем следующий тезис: если Русская идея является квинтэссенцией отечественной культуры, фундаментальным сценарием, поддерживающим ее идентичность, тогда семантическое ядро Идеи должно радикально пересекаться с лексико-семантическим полем существования русской культуры, зафиксированным в Национальном корпусе русского языка. Говоря иными словами, пул наиболее частотных слов из КРИ должен совпадать с набором самых употребляемых слов НКРЯ. Такой подход к рассмотрению Русской идеи не позволяет воссоздать когнитивный сценарий в развернутом виде. Тем не менее первичная категоризация сформированного пула частотных слов КРИ даст возможность обозначить узловые моменты искомого когнитивного сценария. Кроме того, по нашему мнению, подобный подход позволит фундировать результаты предшествующих герменевтических исследований по данной тематике.

### **Материалы и методы**

Использование статистических методов при анализе текстов практикуется исследователями достаточно давно. В настоящее время интерес к методам корпусного анализа

возрос в связи с использованием цифровых инструментов. Существуют различные программы для корпусного анализа текстов, такие как Antconc, Sketch Engine, Voyant Tools и др., имеющие схожий функционал. С их помощью можно получить количественные характеристики текстов: частоту слов и словосочетаний в документах, коллокации слов (наиболее частые совместные употребления слов в документе), коэффициент корреляции их взаимного употребления, также есть возможность изучить контексты анализируемого термина и визуализировать полученные результаты.

Как показал анализ изученной литературы, значительное число корпусных исследований базируется на частотных характеристиках терминов и их коллокаций [Чернышов 2023]. В отдельных исследованиях доказывается, что полученные с помощью цифровых инструментов количественные данные имеют качественную интерпретацию концептуального содержания текста [Яловенко 2022].

В ходе изучения приведенных выше цифровых инструментов нами был проведен их сравнительный анализ, отраженный в таблице 1 по следующим критериям: возможности загрузки своего корпуса текстов; ограничению по числу слов в корпусе; возможным форматам загружаемых документов; возможности работы онлайн; визуализации результатов анализа; встроенным инструментам (кроме инструментов визуализации); тарифам доступа.

**Таблица 1. Сравнение цифровых инструментов для анализа текстов**  
**Table 1. Comparison of Digital Text Analysis Tools**

Критерий	Программа		
	AntConc	Sketch Engine	Voyant Tools
Возможность загрузки своего корпуса	Да	Да	Да
Ограничение по числу слов в корпусе	Нет ограничений	Зависит от тарифа	Нет ограничений
Возможные форматы загружаемых документов	Текстовые файлы (TXT)	Разнообразные форматы (TXT, XML, HTML, PDF и др.)	Разнообразные форматы (TXT, XML, HTML, PDF и др.)
Возможность работы онлайн	Нет (нужна установка на компьютер)	Да	Да
Визуализация результатов анализа	Ограниченные возможности (графики, таблицы)	Широкие возможности (графики, облака слов, диаграммы и др.)	Широкие возможности (графики, облака слов, диаграммы, сети терминов и др.)
Встроенные инструменты (кроме инструментов визуализации)	<ul style="list-style-type: none"> <li>✓ ключевые слова в контексте,</li> <li>✓ частотный анализ,</li> <li>✓ кластеризация слов (n-граммы),</li> <li>✓ анализ коллокаций</li> </ul>	<ul style="list-style-type: none"> <li>✓ ключевые слова в контексте,</li> <li>✓ частотный анализ,</li> <li>✓ кластеризация слов (n-граммы),</li> <li>✓ анализ коллокаций,</li> <li>✓ анализ изменения в частоте использования слов с течением времени</li> <li>✓ сравнение словарных карт двух слов,</li> <li>✓ генерация списка синонимов для слов в корпусе,</li> <li>✓ сложные запросы, включая запросы на естественном языке</li> <li>✓ и др.</li> </ul>	<ul style="list-style-type: none"> <li>✓ ключевые слова в контексте,</li> <li>✓ частотный анализ,</li> <li>✓ кластеризация слов (n-граммы),</li> <li>✓ анализ коллокаций,</li> <li>✓ анализ изменения в частоте использования слов с течением времени</li> <li>✓ анализ корреляции между частотностями различных слов</li> <li>✓ и др.</li> </ul>

Тарифы доступа	Бесплатно	Платные тарифы (есть пробная версия)	Бесплатно (онлайн-версия), платные для отдельных сервисов
----------------	-----------	--------------------------------------	---

Из таблицы 1 видно, что приведенные программы для анализа текста имеют похожие возможности и выбор одной из них зависит от целей исследования. Нами для работы была выбрана программа Voyant Tools [Sinclair, Rockwell 2016] – веб-среда для чтения и анализа текстов с открытым исходным кодом [Voyant Tools Help 2025].

Работа с текстами проходила в несколько этапов:

1. Подбор текстов для исследования.
2. Создание собственного корпуса. На этом этапе тексты были приведены к формату ТХТ, собраны в единый документ, произведена очистка текстов от служебной информации (например, адрес сайта, откуда скачан текст и т. п.).
3. Лемматизация созданного корпуса (преобразование слов корпуса к начальной форме). Для решения данной задачи была использована программа «Яндекса» MyStem – это консольная программа, которая предназначена для определения начальной формы слова и его грамматических характеристик [MyStem ... 2025]. Без этого шага Voyant Tools рассматривает все словоформы термина как отдельные элементы (народ-народ-народу-народом), что снижает качество анализа.
4. Загрузка корпуса в Voyant Tools, удаление не релевантных для целей исследования стоп-слов (частиц, союзов, предлогов, местоимений, вводных слов, числительных) и проведение анализа.

Итак, сбор и первичная обработка данных позволили выявить наиболее употребительные слова как в КРИ, в котором 1 732 413 слов, так и в НКРЯ, в котором 374 449 975 слов.

**Таблица 2. Топ-50 частотных слов по исследуемым корпусам**  
**Table 2. The Top 50 Most Frequently Used Words**

Корпус по Русской идее (КРИ)	Национальный корпус русского языка (НКРЯ)
1) 10 503 = русский = 13	1) 1 084 431 = человек = 2
2) 6 703 = человек = 1	2) 768 251 = время = 6
3) 5 362 = народ = 33	3) 644 576 = знать = 26
4) 4 946 = жизнь = 5	4) 583 032 = дело = 7
5) 4 807 = Россия = 23	5) 531 854 = жизнь = 4
6) 4 112 = время = 2	6) 491 628 = рука = 0
7) 3 251 = дело = 4	7) 431 285 = новый = 11
8) 3 026 = великий = 48	8) 427 548 = слово = 10
9) 2 933 = сила = 19	9) 377 820 = должен = 0
10) 2 783 = слово = 8	10) 369 179 = хотеть = 47
11) 2 702 = новый = 7	11) 321 032 = работа = 0
12) <b>2 664 = история = 0</b>	12) 319 600 = дом = 0
13) 2 504 = мир = 22	13) 315 677 = русский = 1
14) 2 404 = земля = 24	14) 308 641 = друг = 0
15) 2 388 = мысль = 39	15) 293 811 = жить = 0
16) <b>2 311 = князь = 0</b>	16) 292 897 = сделать = 0
17) <b>2 252 = становиться = 0</b>	17) 283 899 = стоять = 0
18) <b>2 100 = исторический = 0</b>	18) 277 879 = город = 0
19) 2 062 = общий = 42	19) 266 305 = сила = 9
20) <b>1 986 = дух = 0</b>	20) 266 075 = казаться = 0



21) <b>1 977 = церковь = 0</b>	21) 250 339 = случай = 0
22) <b>1 907 = идея = 0</b>	22) 248 016 = мир = 13
23) <b>1 895 = государство = 0</b>	23) 240 207 = Россия = 5
24) 1 885 = отношение = 45	24) 234 617 = земля = 14
25) <b>1 880 = Европа = 0</b>	25) 229 080 = ребенок = 0
26) 1 800 = знать = 3	26) 226 789 = Москва = 0
27) <b>1 784 = общество = 0</b>	27) 224 374 = всякий = 0
28) <b>1 783 = религиозный = 0</b>	28) 224 134 = отец = 0
29) 1 784 = образ = 34	29) 221 079 = конец = 0
30) <b>1 696 = духовный = 0</b>	30) 220 231 = прийти = 0
31) <b>1 677 = развитие = 0</b>	31) 219 327 = Бог = 32
32) 1 671 = Бог = 31	32) 212 820 = хороший = 0
33) <b>1 666 = святой = 0</b>	33) 199 775 = народ = 3
34) <b>1 639 = народный = 0</b>	34) 192 971 = образ = 29
35) <b>1 607 = высокий = 0</b>	35) 192 476 = хорошо = 0
36) <b>1 599 = политический = 0</b>	36) 186 393 = женщина = 0
37) <b>1 599 = наука = 0</b>	37) 182 856 = душа = 40
38) <b>1 517 = власть = 0</b>	38) 178 144 = книга = 45
39) <b>1 473 = национальный = 0</b>	39) 178 004 = мысль = 15
40) 1 442 = душа = 37	40) 176 598 = страна = 0
41) <b>1 432 = право = 0</b>	41) 176 357 = письмо = 0
42) <b>1 422 = находить = 0</b>	42) 176 285 = общий = 19
43) <b>1 421 = человеческий = 0</b>	43) 176 271 = молодой = 0
44) <b>1 375 = форма = 0</b>	44) 174 727 = старый = 0
45) 1 359 = книга = 38	45) 174 216 = отношение = 24
46) <b>1 353 = совершенно = 0</b>	46) 172 848 = деньги = 0
47) 1 352 = хотеть = 10	47) 172 123 = работать = 0
48) <b>1 344 = смысл = 0</b>	48) 171 287 = великий = 8
49) <b>1 343 = западный = 0</b>	49) 166 655 = дорога = 0
50) <b>1 315 = характер = 0</b>	50) 158 467 = жена = 0

### Результаты исследования

Приступая к анализу полученных результатов, следует заметить, что, сравнивая 5 самых частотных слов (топ-5) из малого корпуса текстов по Русской идее (25 текстов) из предыдущего исследования [Хамидулин, Бабаева 2024] с 5 самыми частотными словами (топ-5) из большого корпуса текстов по Русской идее (КРИ), мы фиксируем их совпадение. В результате сравнения 50 самых частотных слов (топ-50) из корпуса текстов по Русской идее с 50 самыми частотными словами (топ-50) из НКРЯ обозначились несколько принципиальных моментов.

Во-первых, возникла необходимость выделить из топ-50 частотных слов те, которые встречаются в обоих корпусах, и рассчитать их частотность в процентном значении относительно общего количества слов в каждом корпусе. В результате перед нами предстает следующая картина, отражаемая в таблице 3.

Среди топ-50 22 слова встречаются в обоих корпусах, при этом 20 слов обладают более высокой частотностью в процентном соотношении в КРИ, чем в НКРЯ, и только 2 слова, единственные глаголы, которые встречаются в двух корпусах среди топ-50 («знать» и «хотеть»), оказались в процентном соотношении более частотными для НКРЯ, чем для КРИ. Все из топ-5 в процентном отношении по КРИ оказались более частотными по сравнению с НКРЯ.

**Таблица 3. Результаты частотности топ-22 слов, встречающихся одновременно в КРИ и НКРЯ, относительно общего количества слов в каждом отдельном корпусе (значения приводятся в %)**

**Table 3. The Frequency Ratios of the Top 22 Words that Occur in Both the CRI and the NCRL Relative to the Overall Number of Words in Each Respective Corpus (values are presented in percentages)**

Корпус по Русской идее (КРИ)		Национальный корпус русского языка (НКРЯ)	
1.	Русский – 0,606	1.	Русский – 0,084
2.	Человек – 0,386	2.	Человек – 0,289
3.	Народ – 0,309	3.	Народ – 0,053
4.	Жизнь – 0,285	4.	Жизнь – 0,142
5.	Россия – 0,277	5.	Россия – 0,064
6.	Время – 0,237	6.	Время – 0,205
7.	Дело – 0,187	7.	Дело – 0,155
8.	Великий – 0,174	8.	Великий – 0,045
9.	Сила – 0,169	9.	Сила – 0,071
10.	Слово – 0,160	10.	Слово – 0,114
11.	Новый – 0,155	11.	Новый – 0,115
12.	Мир – 0,144	12.	Мир – 0,066
13.	Земля – 0,138	13.	Земля – 0,062
14.	Мысль – 0,137	14.	Мысль – 0,0475
15.	Общий – 0,119	15.	Общий – 0,0470
16.	Отношение – 0,108	16.	Отношение – 0,046
17.	Знать – 0,103	17.	Знать – 0,172
18.	Образ – 0,102	18.	Образ – 0,051
19.	Бог – 0,096	19.	Бог – 0,058
20.	Душа – 0,083	20.	Душа – 0,048
21.	Книга – 0,078	21.	Книга – 0,047
22.	Хотеть – 0,078	22.	Хотеть – 0,098

Во-вторых, интересной задачей представляется рассмотрение полученных данных с точки зрения сравнения лексической частотности в данных корпусах и устойчивости (базисности) лексики по 100-словному списку Морриса Сводеша [Сводеш 1960] и списку «Лейпциг-Джакарта» Мартина Хаспельмата и Ури Тадмора [Loanwords in the world's languages ... 2009]. В нашем случае результаты выглядят следующим образом: слова в таблице 4 распределены таким образом, что сначала по каждому корпусу в столбце сверху вниз идут те, что совпадают по двум спискам базовой лексики, потом те, что отличаются.

**Таблица 4. Анализ КРИ и НКРЯ в свете списков Сводеша и «Лейпциг-Джакарта»**

**Table 4. The Analysis of the CRI and the NCRL in the Context of the Swadesh and the Leipzig-Jakarta Lists**

Корпус по Русской идее (КРИ)		Национальный корпус русского языка (НКРЯ)	
Список Сводеша	Список «Лейпциг-Джакарта»	Список Сводеша	Список «Лейпциг-Джакарта»
новый (new)	новый (new)	знать (to know)	знать (to know)
знать (to know)	знать (to know)	рука (hand)	рука (arm/hand)
человек (men/person)		новый (new)	новый (new)
земля (earth (предмет или явление природы))		стоять (to stand)	стоять (to stand)
		хороший/хорошо (good)	хороший/хорошо (good)
		прийти/приходить (to come)	прийти/приходить (to come)

Окончание таблицы 4

		человек (men/person)	сделать/делать (to do/to make)
		земля (earth (предмет или явление природы))	дом (house)
		женщина (woman)	ребенок (child (reciprocal of parent))
		дорога (road)	старый (old)

Как замечает С. А. Бурлак, «хотя прямой зависимости между частотностью лексики и ее устойчивостью нет, всё же соотношение между базисностью лексики и ее частотностью является неслучайным» [Бурлак 2021, с. 298]. В ходе нашего сравнительного анализа частотной и устойчивой лексики закономерным выглядит то, что устойчивая лексика встречается чаще (более чем в 2 раза) в топ-50 слов в НКРЯ, чем в КРИ. Только 2 слова из двух списков устойчивых слов входят в оба корпуса текстов: «новый» и «знать». Если же смотреть по списку Сводеша, то еще 2 слова являются общими для топ-50 частотных обоих корпусов: «человек» и «земля», при этом следует помнить, что слово «земля» в корпусе КРИ характеризуется как полисемантическое и варьируется от значения «грунт, почва» до «государство, страна».

Во-третьих, необходимо отметить аспект единства рассматриваемых корпусов (табл. 2). Так, в топ-5 в обоих корпусах входят слова «человек» и «жизнь» с разницей в одну позицию. По частотности в процентном отношении слово «человек» в НКРЯ стоит на первом месте, в то время как в корпусе КРИ это слово занимает второе место, уступив первое место слову «русский», которое в НКРЯ занимает 13-е место по частотности, что представляется вполне соответствующим ожиданиям от КРИ, в котором речь идет о специфике русского национального мировоззрения. Это позволяет сделать вывод о том, что для русского языкового сознания ключевым является жизнь человека, без чего не возможна ни речь, ни философская рефлексия о судьбе России, а весь дискурс Русской идеи вращается вокруг этих двух понятий.

Итак, будем считать слова «человек» и «жизнь» центром семантического поля рассматриваемых концептов. Слова «время» и «дело», входящие в НКРЯ в топ-5, в КРИ входят в топ-7 наряду со словами «слово» (топ-10 по КРИ и топ-8 по НКРЯ) и «новый» (топ-11 по КРИ и топ-7 по НКРЯ), образуя ближнюю периферию объединенного семантического ядра НКРЯ и КРИ в рассматриваемом топ-50 частотных слов. В качестве дальней периферии семантического ядра выделим встречающиеся в обоих корпусах слова, входящие в топ-25, помимо уже названных: русский, Россия, сила, мир, земля (примечательно, что в КРИ они все занимают более частотные позиции по сравнению с НКРЯ, причем отрыв составляет более 10 позиций). В качестве любопытного примера близких параметров частотности отметим два слова: «образ» (топ-29 по КРИ и топ-34 по НКРЯ) и «Бог» (топ-32 по КРИ и топ-31 по НКРЯ); последний случай свидетельствует в пользу того, что для русской языковой картины мира и русского национального самосознания религиозная составляющая, концептуализирующаяся через теологическую категорию «Бог», является одинаково значимой.

В-четвертых, привлекает внимание особенность распределения частотности слов КРИ по сравнению с НКРЯ. Так, среди слов, входящих в топ-50 по частотности в КРИ и не встречающихся в НКРЯ, выделяются следующие: история, князь, становиться, исторический, дух, церковь, идея, государство, Европа, общество, религиозный, духовный, развитие, святой, народный, высокий, политический, наука, власть, национальный, право, находить, человеческий, форма, совершенство, смысл, западный, характер. Эти слова образуют семантическую группу слов с отвлеченными значениями, выражающими момент обобщения, осмысления от наиболее конкретных (князь, Европа), через подгруппу отвлеченных (государство, общество, национальный, народный, западный, человеческий, история, исторический, власть, наука, право) к подгруппе максимально



отвлеченных (идея, смысл, характер, форма, совершенство, высокий, развитие). Выделяется также подгруппа слов, связанных с религиозно-метафизической сферой, которые также ранжируются по уровню отвлеченности (церковь, святой, религиозный, духовный, дух) и свидетельствуют о неотъемлемых для русской национальной языковой картины мира религиозных и духовно-нравственных категорий. Таким образом, для конструирования проекта «Русская идея» используются отвлеченные философские понятия с большим объемом, причем выбор оказывается в пользу понятий с общественно-политическим, философско-историческим и религиозным содержанием.

Данное наблюдение тем более примечательно, что список частотных слов из топ-50 в НКРЯ, не встречающихся в КРИ, носит гораздо более конкретный, предметный характер (рука, работа, дом, друг, книга, страна, город, молодой, старый, деньги, дорога, ребенок, отец, жена, Москва, хотеть, работать и т. д.). Словами, имеющими благодаря своей метафоричности философско-обобщающий потенциал, можно назвать разве что такие, как «жить», «всякий», «хороший», «душа», «мысль», «страна», «отношение», «деньги», «дом», «дорога», что явственно подсвечивает семантическую специфику тезауруса КРИ. Не обходится при этом в КРИ и без привычной нам оппозиции Россия – Европа и соотнесения России с обобщенным западным миром, который выступает в процессе национального самоопределения как традиционный Другой для русской мысли. Неординарным следует охарактеризовать наличие в приводимом нами списке слов, встречающихся в топ-50 КРИ и не встречающихся в топ-50 НКРЯ, двух глаголов «становиться» и «находить» (а также существительного «развитие»), семантика которых указывает на динамический момент совершенствования («развитие»), поиска объективного, уже наличествующего, но постепенно выявляемого элемента («находить») вместе с самосозидающим, формирующим в процессе поиска самого себя, самоконструирующим, приобретающим нечто, усиливающим момент («становиться»), благодаря чему и занимается *русским* государством, обществом, народом и человеком определенное положение в мире и осознание самих себя в качестве самостоятельных, особых, самобытных начал (личностей).

### **Обсуждение**

В итоге данные, представленные в таблицах, позволяют сделать вывод, что семантика КРИ и НКРЯ совпадает на 44 %, т. е. полные совпадения составляют 22 лексемы, что уже частично подтверждает заложенную в основание исследования гипотезу. Самую многочисленную группу слов представляют имена существительные – их 16: человек, народ, жизнь, Россия, время, дело, сила, слово, мир, земля, мысль, отношение, Бог, душа, книга, образ. В структурно-грамматическом аспекте существительные являют собой структуры, обозначающие концепт «вещь», часто соотносимый «в пространстве с физическим телом» [Плотникова 2005, с. 66]. Основной задачей существительных является номинация и категоризация вещей, иначе говоря, имена существительные отвечают за построение каркаса концептуальной картины мира.

Дело в том, что сам процесс категоризации начинается с базового уровня, куда относятся концепты, имеющие одиночный ментальный образ, легко идентифицируемые (см.: [Лакофф, Джонсон 2004, с. 72–73]). Репрезентация концептов данного уровня происходит посредством, если пользоваться терминологией А. Вежбицкой, «семантических примитивов» (см.: [Gladkova, Larina 2018, p. 509]) – универсальных, простых коротких слов, соотносимых с объектами повседневного опыта, на этом основании легко осваиваемых детьми. Сюда относятся такие слова, как человек, рука, книга, стол, собака и прочее. Как видим, большая часть существительных из КРИ не имеет четких внешне воспринимаемых очертаний, вызывает сложности в процессе идентификации, поскольку содержание указанных концептов невыводимо из повседневного опыта и часто требует дополнительных знаний, не говоря уже о необходимости детализации контекста употребления для понимания этих слов. Это справедливо в отношении таких существительных, как дело, время, сила, образ, душа, мысль, Бог, мир. Так, к примеру,

лексема «мир» в русском языке имеет два полисемантических омонима: «Современный большой толковый словарь» у одного омонима фиксирует 9 значений (от «совкупности всех форм материи в земном и космическом пространстве» до «всего, что было создано Богом (по религиозным представлениям)») и 4 значения (от «согласия, отсутствия разногласий, вражды или ссоры» до «покоя, спокойствия») у второго омонима [Современный толковый словарь ... 2008, с. 351].

В процессе детализации картины мира особую роль играют имена прилагательные, основная функция которых заключается, во-первых, в детализации информации о разных уровнях концепта; во-вторых, в фиксации позиции исследователя в отношении объектов окружающего мира. В найденных совпадениях находятся 4 прилагательных. Слово «русский» – относительное прилагательное, которое обозначает признак, присущий объекту (человеку, народу, государству и т. д.) как его внутренняя характеристика, т.е. перед нами основание идентификации объекта, а не актуализированный в момент речи и не приписываемый ему ситуативно признак. Еще три слова: «великий», «новый», «общий» – это качественные прилагательные, обозначающие субъективные характеристики объектов, зависящие от точки зрения познающего субъекта и указывающие на оценку познаваемого.

Таким образом, даже первичный анализ пула существительных и прилагательных дает возможность сделать несколько важных выводов. Во-первых, предметное поле Русской идеи находится за пределами повседневности, поскольку большинство составляющих Идею концептов относится к суперординатному уровню, т. е. составляет высший уровень категоризации. Высокая степень абстракции и семантическая соотнесенность концептов с метафизикой позволяют характеризовать картину мира, построенную указанными концептами, как философскую. Этим обстоятельством объясняется тот факт, что степень пересечения КРИ и НКРЯ составляет менее 50 %: в последнем лексика более предметна и конкретна, поскольку относится к сфере повседневных практик.

Во-вторых, категоризация анализируемых существительных и прилагательных выстраивается по следующим аспектам: *социальный* (народ, русский, Россия, общий, великий (последнее прилагательное характеризует в КРИ существительные «история», «прошлое», «подвиг», «страна», «князь» и пр.)), *темпоральный* (время, новый), *пространственный* (земля (в значении «страна, государство», в КРИ преимущественно используется в значении «ареал русской культуры»)), *гносеологический* (мысль, образ, книга, слово), *метафизический* (Бог, душа, сила (палитра значений от «материального и духовного источника энергии», «способности проявления какого-либо свойства» до «могущества, власти»)).

И, наконец, в-третьих, в набранном пуле слов обнаружено 2 глагола, приводящих «мир в движение», придающих динамику картине мира, зафиксированных в концептах «знать» и «хотеть». Если исходить из макропарадигмы семантической классификации глаголов, где обычно выделяют три семантических поля: «действие», «состояние», «отношение», то оба указанных глагола следует отнести к глаголам «состояния». Таким образом, глаголы из корпуса Русской идеи означают не действие как таковое, а состояние субъекта в отношении выстроенной им картины мира или отдельных его регионов. Глагол «знать» имеет несколько базовых значений: «иметь сведения о чем-либо», «обладать познаниями, представлениями», «быть знакомым» [Современный толковый словарь ... 2008, с. 229]. Глагол «хотеть» имеет два значения: «иметь желание, намерение»; «стремиться к чему-либо» [Там же, с. 909]. В связи с вышесказанным упрощенный когнитивный сценарий обоих глаголов выглядит так: субъект (S) – предикат (P) – объект (O). Если учесть, что из числа существительных четко выделяется семантическая группа «гносеология», то логично предположить, что основополагающим глаголом в корпусе Русской идеи должен быть именно глагол «знать», близкий указанной группе по лексическому значению. При таком содержательном наполнении базовым когнитив-

ным сценарием следовало бы выделить следующий: S (Бог, человек, народ) – P (имеет сведения/обладает знаниями) – об O (Россия, время, мир, человек, Бог, народ и прочее).

### **Выводы**

Представленное исследование подтверждает продуктивность обращения к цифровым инструментам корпусного анализа языковых средств. Данные, полученные с помощью статистических и интеллектуальных методов обработки естественного языка, позволяют, во-первых, охватить и первично обработать огромные объемы информации; во-вторых, получить эмпирическую базу, позволяющую избежать умозрительных выводов и фундировать теоретические тезисы в области исследований культурных смыслов.

Полученные в результате исследования данные в целом подтверждают заявленную во Введении гипотезу: почти половина слов из корпуса Русской идеи и Корпуса национального русского языка совпадают. Это позволяет сделать вывод, что КРИ, действительно, в принципе отражает базовые моменты национальной языковой картины мира. Такое соотношение представляется закономерным, что можно объяснить следующим образом. Во-первых, вся лексика, входящая в КРИ, является составной частью НКРЯ, следовательно, доля высокочастотных слов НКРЯ в пределах рассматриваемого топ-30 фиксирует квинтэссенцию языковой картины мира русского человека. Данный набор слов является одновременно высокочастотным и для КРИ, поскольку последний корпус отражает часть языковой картины мира, связанной с коллективной саморефлексией, без него оказывается невозможным конструирование подобного метанарратива. Во-вторых, большая часть лексики КРИ является специфической в границах рассматриваемых топ-30 слов по причине своеобразия языковых средств, необходимых для осмысления самобытности русского мировоззрения.

Наше исследование позволяет утверждать, что в основе когнитивного сценария русской культуры заложено созерцание метафизических смыслов, зафиксированных в именах существительных с отвлеченным значением. Структурно-семантический анализ корпуса Русской идеи в аспекте пересечения с Национальным корпусом русского языка позволил реконструировать концептосферу, основной интенцией которой является не деятельность как таковая, разворачивающаяся в горизонтальной плоскости повседневности, а созерцательность, имеющая вертикальное устремление, – созерцательность, специфика которой заключается в узнавании предметов, явленных субъекту, формирование представлений о них и трансформация полученной информации в образно-символической форме. Данное положение подтверждается тем фактом, что в совпадающих словах оказались только два глагола. Очевидно, что в фокусе Русской идеи оказывается именно интеллектуальная деятельность, но не практическое преобразование действительности.

### **Список источников**

- Бурлак С. А. Устойчивость и частотность: есть ли корреляция? // Вопросы языкового родства. 2021. № 19/4. С. 293–307.
- Кочеров С. Н., Париллов О. В. Философия русской идеи : монография. М. : ИД «Русская философия», 2022. 300 с.
- Креницкая Г. С. Русская идея : монография. Кн. первая. Томск : Томский гос. ун-т, 2013. 164 с.
- Лакофф Дж., Джонсон М. Метафоры, которыми мы живем / пер. с англ. А. Н. Баранова, А. В. Морозовой ; под ред. и с предисл. А. Н. Баранова. М.: УРСС, 2004. 252 с.
- MyStem: руководство по использованию // Яндекс : сайт. 2025. URL: <https://yandex.ru/dev/mystem/doc/ru/index> (дата обращения: 12.01.2025).
- Межуев В. М. О национальной идее // Вопросы философии. 1997. № 12. С. 3–14.

- Орлов М. О. Русская идея: историко-философские и социальные основания // Известия Саратовского университета. Новая серия. Серия Философия. Психология. Педагогика. 2011. Т. 11, № 3. С. 58–62.
- Плотникова А. М. Когнитивные аспекты изучения семантики (на материале русских глаголов) : учеб. пособие. Екатеринбург : Изд-во Урал. ун-та, 2005. 140 с.
- Сагатовский В. Н. Русская идея: продолжим ли прерванный путь? СПб. : Петрополис, 1994. 217 с.
- Сводеш М. К вопросу о повышении точности в лексикостатистическом датировании // Новое в лингвистике : [сборник]. Вып. 1 / сост., ред. и вступ. ст. В. А. Звегинцева. М.: Изд-во иностранной литературы, 1960. С. 53–87.
- Современный толковый словарь русского языка / [гл. ред. С. А. Кузнецов]. М. : Рипол-Норинт ; СПб. : Рипол-Норинт, 2008. 959 с.
- Сыромятников О. И. Русская идея как принцип поэтики романов Ф. М. Достоевского 1866–1880 годов : автореф. дис. ... д-ра филол. наук : 10.01.01 / Сыромятников Олег Иванович ; [Место защиты: Перм. гос. нац. исслед. ун-т]. Пермь, 2014. 54 с.
- Хамидулин А. М., Бабаева А. В. Опыт апробации контент-анализа в исследовании концепции «Русская идея» // Вестник Мининского университета. 2024. Т. 12, № 1 (46). Статья № 10. DOI 10.26795/2307-1281-2024-12-1-10.
- Чернышов С. А. Дискурсы «Сибирских летописей»: опыт цифрового корпусного анализа текстов // Сибирские исторические исследования. 2023. № 1. С. 128–147. DOI 10.17223/2312461X/39/6.
- Яловенко М. А. К вопросу о применении инструментов для корпусного анализа текстов в рамках когнитивно-стилистического подхода // LXXV Герценовские чтения. Иностранные языки : Междунар. науч. конф. (14–15 апреля 2022 г., Санкт-Петербург) : сборник научных статей / под ред. Т. И. Воронцовой, Н. Ю. Зайцевой. СПб. : Изд-во РГПУ им. А. И. Герцена, 2022. С. 110–115.
- Gladkova A., Larina T. Anna Wierzbicka, Words and the World // Russian Journal of Linguistics. 2018. Vol. 22, no. 3. P. 499–520. DOI 10.22363/2312-9182-2018-22-3-499-520.
- Loanwords in the world's languages : a comparative handbook / ed. by M. Haspelmath, U. Tadmor. Berlin : De Gruyter Mouton, 2009. xxi, 1081 p.
- Sinclair S., Rockwell G. Voyant Tools 2016 : Web-based text analysis suite. URL: <http://voyant-tools.org> (access date: 03.01.2025).
- Voyant Tools Help : website. URL: <https://voyant-tools.org/docs/#!/guide/about> (access date: 03.01.2025).

#### ***Информация об авторах***

**Анастасия Валентиновна Бабаева**, канд. филос. наук, доцент, декан факультета гуманитарных наук, Нижегородский государственный педагогический университет им. Козьмы Минина (Нижний Новгород, Россия).

**Артем Маратович Хамидулин**, канд. филол. наук, доцент кафедры философии и теологии, Нижегородский государственный педагогический университет им. Козьмы Минина (Нижний Новгород, Россия).

**Ольга Валентиновна Смышляева**, старший преподаватель кафедры прикладной информатики и информационных технологий в образовании, Нижегородский государственный педагогический университет им. Козьмы Минина (Нижний Новгород, Россия).

#### ***Information about the authors***

**Anastasia V. Babaeva**, Cand. Sci. (Philosophy), Assoc. Prof., Dean of the Faculty of the Humanities, Minin Nizhny Novgorod State Pedagogical University (Minin University) (Nizhny Novgorod, Russia).

**Artem M. Khamidulin**, Cand. Sci. (Philology), Assoc. Prof. at the Department of Philosophy and Theology, Minin Nizhny Novgorod State Pedagogical University (Minin University) (Nizhny Novgorod, Russia).

**Olga V. Smyshliaeva**, Senior lecturer at the Department of Applied Informatics and Information Technologies in Education, Minin Nizhny Novgorod State Pedagogical University (Minin University) (Nizhny Novgorod, Russia).

*Статья поступила в редакцию | Submitted 08.04.2025.*

*Одобрена после рецензирования | Revised 24.05.2025.*

*Принята к публикации | Accepted 25.05.2025.*