

Научная статья

УДК 81-13

DOI: <https://doi.org/10.18721/JHSS.15303>



## ЛИНГВОПРАГМАТИЧЕСКИЙ АНАЛИЗ ГЛАГОЛОВ СТАНОВЛЕНИЯ ЦВЕТА (НА ОСНОВЕ ТЕХНОЛОГИЙ НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА)

**К.Ф. Васильев** 

Московский городской педагогический университет,  
Москва, Российская Федерация

✉ [vasilevkv@mgpu.ru](mailto:vasilevkv@mgpu.ru)

**Аннотация.** Статья анализирует возможности поисковых и аналитических технологий Национального корпуса русского языка (НКРЯ). При рассмотрении лексической группы с устойчивыми семантико-грамматическими признаками исследователь не всегда обладает возможностью неограниченного поиска корпусного материала. Лексико-грамматический поиск НКРЯ не позволяет осуществить отбор глаголов по семантическому признаку «цвет», однако рассматриваемые в исследовании инструменты, например «Портрет слова», позволяют частично компенсировать этот пробел. Сервис может быть использован для отбора лексики, появляющейся в однородных контекстах, установления её семантической близости, определения семантико-грамматических признаков, которые могут лечь в основу усложнённого запроса в корпус. В результате исследования получена эмпирическая база для анализа глаголов становления цвета: определены ассоциативные поля для выбранных глаголов, их семантическая близость и сочетаемость на основе коллокатов, выделены грамматические и семантические признаки, по которым может корректироваться поиск единиц данной группы лексики в НКРЯ. Выявлено, что при разметке НКРЯ выделяется больше критериев, чем это доступно в лексико-грамматическом поиске, что позволяет наметить вектор дальнейшего развития корпуса.

**Ключевые слова:** глаголы становления цвета, корпусное исследование, семантика, токенизация, национальный корпус русского языка.

**Для цитирования:** Васильев К.Ф. Лингвопрагматический анализ глаголов становления цвета (на основе технологий Национального корпуса русского языка) // Terra Linguistica. 2024. Т. 15. № 3. С. 28–34. DOI: 10.18721/JHSS.15303



## VERBS DENOTING CHANGES OF COLOUR: LINGUOPRAGMATIC ANALYSIS BASED ON RUSSIAN NATIONAL CORPUS

K.F. Vasilev 

Moscow City University, Moscow, Russian Federation

 [vasilevkv@mgpu.ru](mailto:vasilevkv@mgpu.ru)

**Abstract.** The paper looks into the searching potential of Russian National Corpus (RNC). Investigating the lexical group with constant semantical and grammatical features, researchers are to rely on corpus search technologies, the corpus offering rather limited application. Lemmas and tags of RNC search don't include colour as a search tag for verbs so this function is passed to the instruments referred to in the research suggested, for example, 'Word at glance'. The service can be used to select lexical units featuring in homogeneous contexts, to establish their semantic proximity, and to determine semantical and grammatical characteristics applicable for the follow-up corpus inquiry. The study focuses on the verbs denoting changes of colour, their associative field, semantic proximity and combinability in collocations, grammatical and semantic features, suitable for regulating subsequent corpus queries. The findings clearly indicate that RNC is hardwired to recognize more criteria at the markup stage than those available for the lexical and grammatical search, and this outlines the perspectives for further corpus development.

**Keywords:** verbs denoting changes of colour, corpus study, semantics, tokenization, Russian National Corpus.

**Citation:** Vasilev K.F., Verbs denoting changes of colour: linguopragmatic analysis based on Russian National Corpus, *Terra Linguistica*, 15 (3) (2024) 28–34. DOI: 10.18721/JHSS.15303

### Введение

Цвет и цветообозначения в качестве объекта лингвистического исследования находятся в сфере внимания отечественных и зарубежных учёных с 60-х годов XX века. Данная группа лексики исследовалась: в рамках различных подходов в русле лексикостатистики и лексикологии, что позволяет определить формальные критерии выделения цветообозначений [1]; в рамках сравнительно-исторического языкознания, когда исследовалось диахроническое развитие как отдельных цветообозначений, так и комплексное формирование этой группы лексики в русском языке [2, 3]; А. Вежицкая, С. В. Ионова предприняли концептуальный анализ цветообозначений, осмысляя их семантику [4, 5]. При этом при исследовании цветообозначений А. Вежицкая сделала попытку соотнести их с цветом эталонных для данной культуры объектов [6], используя остенсивный метод. Цвет как тема исследования превосходит по объёму посвящённых ему работ многие другие темы (см., например, обзоры в [7, 8]). Зарубежные исследования цвета появились на почве лингвистической антропологии и типологии в работах Б. Берлина и П. Кея [6], а современные зарубежные лингвистические исследования используют междисциплинарный комплексный подход к рассмотрению явления цвета. Так, например, в междисциплинарной перспективе рассматриваются восприятие, осмысление и интерпретация цвета сразу в нескольких областях знания, в том числе и в лингвистике [9]. Высокий интерес к теме цвета и цветообозначений обуславливает появление новых исследований с применением современных подходов, а именно цифровых методов.

Цифровые методы позволили разработать специальные, собственно-цифровые методы обработки языкового материала [10], которые также могут быть использованы при создании



эмпирической базы исследования [11, 12], при обработке больших объёмов данных в исследовательских целях [13], в том числе для исследования контекстов в социальных практиках, см. обзор в [14]. Это, собственно, и привело к формированию специального направления корпусных исследований. Главным представительным ресурсом в отечественной лингвистике по поиску языковых данных можно назвать Национальный корпус русского языка (НКРЯ)<sup>1</sup>.

Основной корпус НКРЯ содержит 375 миллионов словоупотреблений, которые сопровождаются лингвистической разметкой вкупе с удобными инструментами поиска и анализа, что существенно облегчает работу исследователей. Помимо привычных инструментов, с 2023 года в корпусе представлен сервис «Портрет слова»<sup>2</sup>, включающий новые модели автоматической разметки НейроКРЯ. Вместе с тем оснащённость корпуса не позволяет в полной мере реализовать все потребности современных исследователей. Продемонстрируем это на примере создания эмпирической базы корпусного лингвистического исследования, рассматривающего глаголы, содержащие сему *цвет*.

Находясь на периферии концепта *цвет*, глаголы с семой *цвет* привлекают тем не менее внимание исследователей: например, предметом исследования уже служила сочетаемость данных единиц [15], однако приходится признать, что не все особенности семантики данных глаголов получили исчерпывающее описание, а именно — интерпретация глаголов репрезентации цвета как глаголы бытийности и их коннотативные особенности.

### Методика исследования

При проведении лингвистического исследования возникает необходимость привлечения большого массива лингвистических данных, содержащих искомые единицы.

Среди представленных средств поиска оптимальным при поиске лексических единиц по заданным параметрам является лексико-грамматический поиск в основного корпуса НКРЯ. В настоящее время в корпусе реализована возможность внести в условия поиска грамматический признак (глагол), однако поиск по семантическому признаку (цвет) среди глаголов не представлен. При этом в таксономии семантических признаков глагола существует критерий *свет*, который позволяет с помощью лексико-грамматического поиска отобрать все вхождения глаголов с семой *свет*. Вместе с тем в лингвистических исследованиях группы колоронимов с семами *цвет* и *свет* часто не разграничиваются. То есть условия поиска НКРЯ обуславливают разграничение этих лексических групп уже на этапе создания эмпирической базы исследования.

Итак, невозможно корпусными методами выделить глаголы с семой *цвет* на основе грамматико-семантических критериев поиска, запрос поиска необходимо сужать. Решению данной исследовательской задачи может способствовать инструмент «Портрет слова» с внедрёнными технологиями машинной обработки естественного языка НейроКРЯ.

Сервис «Портрет слова» позволяет запросить информацию о слове, вводимом в словарной форме, лемме. Информация о слове отображается при помощи встроенных в интерфейс корпуса приложений-виджетов, среди которых: «Скетчи» (коллокации со словами разных частей речи), «О слове» (грамматические, семантические и словообразовательные признаки слова), «Частотность слова», «Морфемный разбор», «Однокоренные слова» (в том числе генерируемые нейронной сетью), «Похожие слова» (ассоциаты, полностью генерируемые нейронной сетью), «Статистика текстов» (диаграмма, составленная согласно статистике по набору метаатрибутов, например полу автора или стилю), «Распределение по годам» и «Примеры». Рассмотрим, к примеру, результаты выдачи виджетов, полученные при введении леммы *белеть*.

### Результаты

В рамках предпринятого исследования на основе коллокаций со словами разных частей речи в виджете «Скетчи» выделены глаголы, появляющиеся в однородных контекстах, например,

<sup>1</sup> Национальный корпус русского языка. URL: <https://ruscorpora.ru/> (дата обращения: 01.09.2023).

<sup>2</sup> Сервис НКРЯ «Портрет слова». URL: <https://ruscorpora.ru/word/main> (дата обращения: 04.10.2023).



Скетчи ?

белеть Глагол

	Косвенные дополнения		Предложные дополнения		Сочинные глаголы	
9,11	1. рубашка	9,64	1. темнота	7,57	1. чернеть	8,21
8,06	2. известь	8,74	2. зелень	7,56	2. желтеть	7,25
7,39	3. известка	8,22	3. сумрак	7,51	3. сереть	7,11
6,92	4. сруб	8,05	4. полумрак	7,43	4. синеть	6,98
6,83	5. парус	8,03	5. сумерки	7,02	5. зеленеть	6,95
6,67	6. палатка	7,89	6. луна	6,76	6. белеть	6,87
6,67	7. кость	7,85	7. полутьма	6,71	7. золотиться	6,77
6,65	8. барашек	7,84	8. фон	6,65	8. краснеть	6,71
6,59	9. брохо	7,82	9. мрак	6,59	9. пестрить	6,45
6,51	10. белизна	7,78	10. иной	6,2	10. мерцать	6,3

Рис. 1. Коллокации глагола белеть в сервисе «Портрет слова»  
 Fig. 1. Collocations of the verb whiten in the service “Portrait of the word”

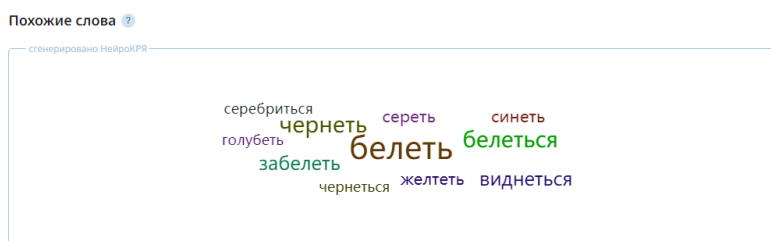


Рис. 2. Семантические ассоциаты к глаголу белеть в сервисе «Портрет слова»  
 Fig. 2. Semantic associations to the verb whiten in the service “Portrait of the word”

*чернеть*, *желтеть*, что гипотетически позволило отобрать и другие глаголы данной группы. Однако среди глаголов, часто встречающихся в однотипных позициях, встречаются и такие, как *пестрить*, *мерцать* (рис. 1), что подчёркивает тесную связь глаголов с семантикой цвет и свет, но ограничивает данный метод как способ отбора единиц одной семантической группы.

Обращение к виджету «О слове» позволило получить характеристику глаголу *белеть* с точки зрения грамматики и семантики. Технология анализа языковых единиц выделяет грамматические (*глагол, непереходный, несовершенный*) и семантические (*изменение состояния или признака; некаузативные глаголы; отадъективные; цвет*) категории слова *белеть*. Технология поиска и разметки позволяет осуществить запрос в основной корпус по всем выделенным грамматическим категориям и лишь по двум семантическим (*изменение состояния или признака; некаузативные глаголы*).

Виджет «Похожие слова» предлагает 10 семантических ассоциатов к глаголу *белеть* (рис. 2). Схема имеет вид облака слов, в котором центральное место занимает искомая единица, а вокруг машинными методами выстраиваются слова, которые оказывались в наиболее схожих контекстах. Семантическая близость указывается в виде коэффициента при наведении мыши на слово и рассчитывается на основе моделей дистрибутивной семантики.

Сгенерированные ассоциаты позволяют эмпирически проверить близость исследуемых единиц. На примере глагола *белеть* благодаря данным виджета находим подтверждение дихотомии *белеть-чернеть*, противопоставление ахроматического *белеть* хроматическим *синеть*, *голубеть*, *желтеть*. Обнаруживаем, что в некоторых контекстах *белеть* обладает семантикой бытийности и приближен к глаголу *виднеться*.

Данный инструмент может использоваться при наборе материала методом сплошной выборки, однако из-за того, что выделение ассоциатов происходит полностью автоматически, в списках

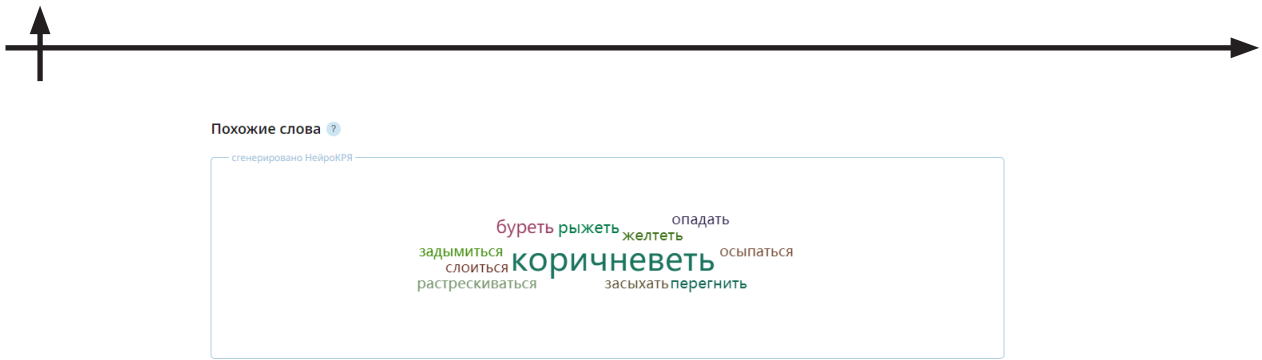


Рис. 3. Семантические ассоциаты к глаголу коричневеть в сервисе «Портрет слова»

Fig. 3. Semantic associations to the verb to brown in the service “Portrait of the word”

могут встречаться ошибки, например неправильно образованные слова, интуитивно не вполне понятные сопоставления слов (см. рис. 3).

Данные прочих виджетов сервиса «Портрет слова» в контексте данного исследования представляют меньшую значимость.

Встроенные технологии обработки естественного языка, отвечающие за анализ единиц в сервисе «Портрет слова», справляются с большим объёмом данных корпуса и выделяют множество характеристик слов, не привязанных к разметке корпуса, в то время как существующие в настоящее время инструменты поиска имеют ограниченный потенциал ввиду зависимости от разметки. Так, в предложенном примере приведена иллюстрация, как некоторые семантические характеристики слов типа цвет в группе глаголов выделяются технологиями анализа, но не внедряются в разметку корпуса, что требует дальнейшей оптимизации главного отечественного корпуса русского языка.

### Выводы

Таким образом, исследование позволило создать эмпирическую базу для анализа глаголов становления цвета: выявлены ассоциативные поля для выбранных глаголов, определена их семантическая близость и сочетаемость на основе коллокатов, выделены грамматические и семантические признаки глаголов, по которым может корректироваться поиск единиц данной группы лексики в НКРЯ. Данный корпус как массивный исследовательский ресурс, позволяющий анализировать большие объёмы языковых данных с помощью высокотехнологичных инструментов поиска и подсчёта, в целом удовлетворяет запрос современных исследователей. Вместе с тем потенциал узкоспециализированных исследований ограничен технологией поиска, которой только предстоит реализовать многие возможности, например, расширить поиск по семантическим параметрам, что стало препятствием при создании эмпирической базы в данном исследовании. Так, в процессе токенизации, то есть присвоения определённых атрибутов языковым единицам в широком контексте, поисковые возможности не включают токен «цвет» в блоке критериев семантического поиска среди глаголов, что не позволяет отобрать все вхождения глаголов с семой *цвет* в корпус. Однако появление новых сервисов, например «Портрет слова», и внедрение нейросетевых технологий вносят значительный вклад в упрощение работы с корпусом, в том числе для узкоспециализированных исследований.

### СПИСОК ИСТОЧНИКОВ

1. **Москович В.А.** Статистика и семантика: Опыт статистического анализа семантического поля. М.: Наука. 1969. 304 с.
2. **Бахилина Н.Б.** История цветообозначений в русском языке. М.: Наука, 1975. 288 с.



3. **Василевич А.П., Кузнецова С.Н., Мищенко С.С.** Цвет и названия цвета в русском языке. М.: КомКнига, 2005. 216 с.
4. **Вежбицкая А.** Язык. Культура. Познание: пер. с англ. / отв. ред. М.А. Кронгауз; вступ. ст. Е.В. Падучевой. М.: Русские словари, 1996. 416 с.
5. **Ионова С.В.** Аспекты исследования письменного текста как объекта лингвистической экспертизы // Вестник Волгоградского государственного университета. Серия 2: Языкознание. 2017. Т. 16, № 2. С. 28–38. DOI: 10.15688/jvolsu.2.2017.2.3
6. **Кожкарева Н.Б.** Сколько цветов в радуге? // Наука из первых рук. 2018. № 2–3 (78). С. 90–99.
7. **Сластникова Т.В., Черкашина Е.И.** Цвет и цветообозначение в лингвистических исследованиях. М.: Языки Народов Мира, 2021. 240 с.
8. **Большова А.В.** Понятие «цвет» в современной отечественной и зарубежной лингвистике // Наука в мегаполисе. 2022. № 5 (40).
9. *Anthropology of Color: Interdisciplinary multilevel modeling* / ed. by R.E. MacLaury, G.V. Paramei, D. Dedrick. John Benjamins Publishing, 2007. 485 p. DOI: 10.1075/z.137
10. Прикладная и компьютерная лингвистика: коллективная монография / под ред. И.С. Николаева, О.В. Митрениной, Т.М. Ландо. М.: URSS, 2016. 320 с.
11. **Фролова Е.А.** Исследование современного состояния конструкции с расщепленным инфинитивом на базе корпусов английского языка // В многомерном пространстве современной лингвистики: сборник работ молодых ученых / отв. ред. О.А. Сулейманова; сост. К.С. Карданова-Бирюкова. Т. 2. М.: ПРИНТИКА, 2021. С. 245–248.
12. **Шестопалова А.М.** Работа с поисковыми цифровыми ресурсами при анализе синтаксических структур // Лингвистика, перевод, межкультурная коммуникация: материалы XXIII научно-практической конференции 22 ноября 2021 г. Екатеринбург: Универсальная типография «Альфа Принт», 2022. С. 243–250.
13. **Сулейманова О.А.** Семантический эксперимент: новые возможности в системе координат big data // Когнитивные исследования языка. 2019. № 36. С. 427–432.
14. **Чернявская В.Е.** Текст и социальный контекст: Социолингвистический и дискурсивный анализ смыслопорождения. М.: URSS, 2021. 208 с.
15. **Аль-Хаснави А.Р.М.** К вопросу о сочетаемости глаголов становления и проявления цветного признака с разными лексико-семантическими группами имен существительных // Филологические науки. Вопросы теории и практики. 2017. № 5–2 (71). С. 47–52.

## REFERENCES

- [1] **Moskovich V.A.**, Statistika i semantika: Opyt statisticheskogo analiza semanticheskogo polya [Statistics and semantics: Experience in statistical analysis of the semantic field], Nauka, Moscow, 1969.
- [2] **Bakhilina N.B.**, Istoriya tsvetooboznacheniy v russkom yazyke [History of color terms in Russian], Nauka, Moscow, 1975.
- [3] **Vasilevich A.P., Kuznetsova S.N., Mishchenko S.S.**, Tsvet i nazvaniya tsveta v russkom yazyke [Color and color names in Russian], KomKniga, Moscow, 2005.
- [4] **Wierzbicka A.**, Semantics, Culture and Cognition: transl. from English], Resp. ed. M.A. Krongauz, Introd. by Ye.V. Paducheva, Russkiye slovari, Moscow, 1996.
- [5] **Ionova S.V.**, Aspects of Research Written Text as an Object of Forencis Linguistic Expertise, Science Journal of Volgograd State University. Linguistics, 16 (2) (2017) 28–38. DOI: 10.15688/jvolsu.2.2017.2.3
- [6] **Koshkareva N.B.**, Skolko tsvetov v raduge? [How many colors are there in a rainbow?], Science First Hand, 2–3 (78) (2018) 90–99.
- [7] **Slastnikova T.V., Cherkashina Ye.I.**, Tsvet i tsvetooboznacheniy v lingvisticheskikh issledovaniyakh [Color and color designation in linguistic research], Yazyki Narodov Mira, Moscow, 2021.
- [8] **Bolshova A.V.**, The concept of color in modern Russian and foreign linguistics, Science in a Megapolis, 5 (40) (2022).
- [9] *Anthropology of Color: Interdisciplinary multilevel modeling*, Ed. by R.E. MacLaury, G.V. Paramei, D. Dedrick, John Benjamins Publishing, 2007.
- [10] *Prikladnaya i kompyuternaya lingvistika: kollektivnaya monografiya* [Applied and computational linguistics: collective monograph], Ed. by I.S. Nikolayev, O.V. Mitrenina, T.M. Lando, URSS, Moscow, 2016.



[11] **Frolova Ye.A.**, Issledovaniye sovremennogo sostoyaniya konstruktсии s rasshcheplennym infinitivom na baze korpusov angliyskogo yazyka [Study of the current state of the construction with a split infinitive based on English corpora], V mnogomernom prostranstve sovremennoy lingvistiki: sbornik rabot molodykh uchenykh [In the multidimensional space of modern linguistics: a collection of works by young scientists], Resp. ed. O.A. Suleymanova, Comp. K.S. Kardanova-Biryukova, 2, PRINTIKA, Moscow, 2021.

[12] **Shestopalova A.M.**, Rabota s poiskovymi tsifrovymi resursami pri analize sintaksicheskikh struktur [Working with searchable digital resources when analyzing syntactic structures], Lingvistika, perevod, mezhkulturnaya kommunikatsiya: materialy XXIII nauchno-prakticheskoy konferentsii 22 noyabrya 2021 g. [Linguistics, translation, intercultural communication: materials of the 23d scientific and practical conference, November 22, 2021], Universalnaya tipografiya "Alfa Print", Yekaterinburg, 2022.

[13] **Suleymanova O.A.**, Semanticheskiy eksperiment: novyye vozmozhnosti v sisteme koordinat big data [Semantic experiment: new possibilities in the big data coordinate system], Cognitive Studies of Language, 36 (2019) 427–432.

[14] **Chernyavskaya V.E.**, Text and Its Social Context, URSS, Moscow, 2021.

[15] **Al-Hasnawi A.R.M.**, On Compatibility of the Verbs of Formation and Manifestation of Colour Feature with Different Lexico-Semantic Groups of Nouns, Philology. Theory & Practice, 5–2 (71) (2017) 47–52.

#### **СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR**

**Васильев Кирилл Фирдовсиевич**

**Kirill F. Vasilev**

E-mail: [vasilevkv@mgpu.ru](mailto:vasilevkv@mgpu.ru)

<https://orcid.org/0000-0002-4033-4211>

*Поступила: 29.05.2024; Одобрена: 17.09.2024; Принята: 02.10.2024.*

*Submitted: 29.05.2024; Approved: 17.09.2024; Accepted: 02.10.2024.*