

Частотное поведение прилагательных цвета в русских поэтических текстах

А. Ц. Масевич¹, В. П. Захаров²

¹ Санкт-Петербургский государственный институт культуры

² Санкт-Петербургский государственный университет
Санкт-Петербург, Россия

Аннотация

Лингвистические корпуса и компьютерные технологии дают возможность осуществлять исследования и получать результаты на больших текстовых массивах, что прежде, до появления этих средств, было недостижимо. Настоящая статья описывает некоторые количественные данные использования прилагательных, обозначающих цвета, в русских поэтических текстах. Произведено сравнение частотности базовых прилагательных цвета (белый, черный, красный, зеленый, желтый, синий, голубой, коричневый, оранжевый, розовый, фиолетовый) в текстах четырех корпусов НКРЯ – основного, газетного, устного и поэтического. Сравнение осуществлялось двумя способами: в каждом корпусе определялась относительная частота (ipm) каждого из перечисленных выше прилагательных, а также ipm всех прилагательных, имеющих тег «цвет» семантической разметки НКРЯ. Оба способа определения показали, что в текстах поэтического корпуса частотность прилагательных цвета значительно выше, чем в текстах трех других корпусов (основном, газетном и устном). Установлено также, что убывание значений частоты прилагательных цвета в текстах всех четырех корпусов в значительной степени коррелирует с эволюционно-исторической схемой Берлина – Кея. Помимо этого, проведено исследование частоты встречаемости прилагательных цвета в поисковой системе Государственного Эрмитажа (массив аннотаций к произведениям искусства), где нам пришлось ограничиться абсолютным числом текстов, содержащих то или иное прилагательное цвета, поскольку данная поисковая система не содержит лингвистических характеристик. Корреляция полученных данных со схемой Берлина – Кея выражена в данном случае намного слабее, чем в корпусах НКРЯ. В статье детально описана методика извлечения прилагательных цвета из авторских подкорпусов поэтического корпуса НКРЯ и создания списков этих прилагательных средствами менеджера Sketch Engine из текстов авторов, отсутствующих в поэтическом корпусе НКРЯ. На основе анализа 180 лемм прилагательных цвета, извлеченных из текстов тридцати шести авторов, произведена категоризация прилагательных цвета. Для анализа и категоризации привлечены также и данные поисковой системы Эрмитажа. Приведены также количественные данные о представленности (репрезентации) выделенных категорий прилагательных в исследованных текстах. В ходе исследования выявлены некоторые недостатки семантической разметки в НКРЯ. Таким образом, мы выявили некоторые закономерности частотного поведения прилагательных цвета в русском языке и предложили один из возможных вариантов их классификации.

Ключевые слова

прилагательные цвета, теория базовых цветовых терминов Берлина – Кея, Национальный корпус русского языка, поэтический корпус, семантическая разметка

Для цитирования

Масевич А. Ц., Захаров В. П. Частотное поведение прилагательных цвета в русских поэтических текстах // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 21–48. DOI 10.25205/1818-7935-2019-17-1-21-48

Frequency Behavior of Color Adjectives in Russian Poetic Texts

Andrey Ts. Masevich¹, Victor P. Zakharov²

¹ St. Petersburg State Institute of Culture

² St. Petersburg State University
St. Petersburg, Russian Federation

Abstract

Linguistic corpora and computer technologies make it possible to search and analyze large amount of unstructured texts. This paper describes in detail the method we used to extract adjectives of color from the poetic texts found in the Poetry corpus of the Russian National Corpus and from various Internet sources. Using a base of 180 lexical units extracted from the poetic texts of 36 authors, we devised a categorization scheme for adjectives of color; this scheme al-

so incorporated data obtained from the Hermitage Museum information system. It includes five categories, with the largest (“Derivatives”) broken down into three subcategories. Our paper further provides quantitative data indicating the extent to which the elements of the different categories are represented in the texts; and from this data we were able to draw a preliminary conclusion vis-à-vis the use of adjectives of color by various authors. Specifically, we compared the frequency of use of basic adjectives of color (белый – white; чёрный – black; красный – red; зелёный – green; жёлтый – yellow; синий – blue; голубой – light blue; коричневый – brown; оранжевый – orange; розовый – pink; and фиолетовый – violet) in the texts of four corpora of the Russian National Corpus, i.e., the Basic corpus, Newspaper corpus, Oral corpus, and Poetry corpus.

Our paper describes some patterns of frequency behavior for adjectives denoting colors in Russian poetic texts. We arranged the adjectives retrieved from each corpus in order of decreasing frequency and found that, in all four corpora, the order of the adjectives of color largely correlates with the evolutionary theory of Berlin – Kay which describes the order of appearance of color adjectives in the historical development of different languages. The comparison showed that the frequency of adjectives of color in the Poetry corpus is significantly higher than in the other three corpora. In addition, we searched the information system of the State Hermitage Museum and established that the frequency correlation between adjectives of color and the Berlin – Kay evolutionary model is expressed there weaker than in the RNC corpora. Also, in the course of our study, we found a few semantic tagging errors in the Russian National Corpus. The patterns of frequency behavior of color adjectives revealed in the Russian language may become a reliable basis for further research. Their classification needs more attention too.

Keywords

color adjectives, Berlin – Kay’s basic color terms theory, National corpus of Russian Language, poetic corpus, semantic tagging

For citation

Masevich Andrey Ts., Zakharov Victor P. Frequency Behavior of Color Adjectives in Russian Poetic Texts. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 21–48. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-21-48

Введение

Обозначение цвета – исключительно популярная тема в лингвистике. В данной работе мы попытаемся ввести в рассмотрение пять принципиально новых аспектов проблемы.

1. Аспект лингвистической статистики. Все наши выводы базируются на количественных результатах, полученных на представительных корпусах.

2. Эволюционный аспект. Выявлена частичная корреляция между эволюционной теорией базовых цветовых терминов Брента Берлина и Пола Кея (Brent Berlin & Paul Kay) [Berlin, Kay, 1969] и статистикой частоты встречаемости прилагательных цвета в русском языке.

3. Предложен вариант классификации прилагательных цвета и показаны отличия в использовании прилагательных цвета разных типов.

4. Выявлены особенности использования прилагательных цвета в текстах различных жанров и в языке для специальных целей, а именно в описаниях произведений изобразительного искусства.

5. Описано частотное поведение единиц данной лексической группы на материале поэтических текстов.

В качестве материала исследования использовались:

1) поэтический корпус Национального корпуса русского языка (НКРЯ) [Гришина и др., 2009];

2) корпуса поэтических текстов собственной генерации, созданные специально для целей данного исследования в системе Sketch Engine;

3) основной, газетный и устный корпуса НКРЯ;

4) база данных описаний коллекций Государственного Эрмитажа.

В качестве инструментов исследования использовались:

1) поисковый аппарат (корпусный менеджер) НКРЯ, в том числе поиск по семантическому признаку;

2) корпусный менеджер Sketch Engine;

3) информационно-поисковая система Государственного Эрмитажа;

4) электронные таблицы MS Excel 2010.

Лингвистический корпус является моделью естественного языка. Очевидно, что, для того чтобы модель статистически достоверно отражала явления естественного языка, корпус должен быть репрезентативным (достаточно большим по объему) и сбалансированным (отражать различные литературные жанры и функциональные стили языка, причем в пропорции, которая соответствует доле данного жанра или стиля в реальном массиве текстов). Для русского языка больше всего этим требованиям удовлетворяет основной корпус НКРЯ (283 млн словоупотреблений). Исследование тех или иных лингвистических единиц или явлений применительно к стилевому многообразию требует, в свою очередь, корпусов определенных жанров (регистров). НКРЯ дает для таких исследований достаточный набор специальных корпусов. В нашем исследовании мы ограничили свой выбор такими специфическими подъязыками, как газетный язык, поэтический язык и устная речь, которым в НКРЯ соответствуют корпусы достаточного объема (228 521 421, 10 967 173 и 12 113 491 словоупотреблений, соответственно). Исследование проводилось в сравнении данных этих корпусов с данными основного корпуса.

Поскольку перечисленные корпуса содержат различное количество словоупотреблений, то абсолютные показатели количества тех или других единиц нельзя использовать как критерий для вывода о большей или меньшей частотности слова в одном корпусе по сравнению с другим. Поэтому для сравнений, как правило, используется относительная частота (частное от деления абсолютной частоты на объем соответствующего корпуса), которая «нормируется» путем преобразования количества случаев встречаемости слова по стандартной шкале, обычно в пересчете на 1 млн слов – *instances per million (ipm)* (относительная частота, умноженная на миллион). Нормированная частота обеспечивает основание для сравнения частотности каких-либо единиц в разных корпусах.

Одна из трудностей при анализе лексики в корпусной лингвистике – это проблема многозначности. И в нашем случае, нет, пожалуй, ни одного прилагательного цвета, по крайней мере из числа принадлежащих к основным категориям, которое не употреблялось бы в нескольких значениях, в том числе в переносном смысле – метафорически либо символически. Достаточно вспомнить выражения «белый свет», «чёрный список», «красной нитью», «оранжевая революция», «желтая пресса», «тоска зелёная», «синий чулок», «коричневая чума». Выскажем предположение: устойчивые словосочетания с изменением значения слова, как правило, образуют прилагательные, означающие базовые категории цвета, и, что также интересно, эти сочетания часто имеют негативную коннотацию.

Проблему учета многозначности можно было бы решить при наличии семантических корпусов с разметкой по значениям, однако таких корпусов мы не знаем. «Ручной» анализ при обработке больших объемов выходных данных нереален. Поэтому в данном исследовании мы попытались подключить к исследованию базу данных, где прилагательные цвета, как нам кажется, в подавляющем большинстве случаев используются в их прямом цветовом значении. Это база данных описаний предметов коллекции Государственного Эрмитажа. Однако поисковая система Эрмитажа на лексический запрос выдает только абсолютное число документов (описаний предметов коллекций), в которых искомая лексическая единица встретилась. Остается неизвестным суммарное число словоупотреблений в описаниях, образующих базу данных, из-за чего эти результаты поиска оказываются не сопоставимыми с характеристиками, получаемыми на основе корпусов НКРЯ и Sketch Engine. Тем не менее, сами по себе полученные результаты представляют интерес.

Одной из замечательных особенностей НКРЯ, отличающей его от многих других корпусов, является наличие семантической разметки [Кустова и др., 2005]. Ее назначение – выделение (поддержка поиска) групп лексических единиц, объединенных общими семантическими признаками. Семантические признаки приписываются всем словоформам знаменательных слов корпуса. К сожалению, в ходе нашей работы при использовании тегов семантической разметки в ней выявлено значительное число ошибок. Тем не менее, мы считаем,

что, несмотря на несовершенство существующей разметки, использование ее позволяет получить определенные результаты, которые нельзя получить никаким другим способом.

Корреляция между эволюционной теорией Берлина и Кея и частотой встречаемости прилагательных цвета в корпусах русского языка

Теория базовых цветовых терминов Берлина и Кея

Теория базовых цветовых терминов (категорий) Берлина – Кея состоит в том, что языки мира содержат полностью или частично общий набор слов цветообозначений, и эти слова появляются в языках в определенном порядке [Berlin, Kay, 1969]. В английском языке авторы выделяют 11 таких прилагательных, в славянских, в частности в русском, – 12.

Критерии принадлежности к базовым категориям следующие:

- 1) слово должно быть неприменомным и не относиться к сложным словам;
- 2) его значение не должно быть уже значения другого имени цвета, указывающего на какой-либо близкий оттенок;
- 3) слово должно обладать широкой сочетаемостью;
- 4) для носителей данного языка слово должно быть психологически выделенным, значимым (salient).

Для английского языка Берлин и Кей указывают на цветовые термины, представленные на рис. 1. Для русского языка предлагается еще двенадцатое прилагательное – «голубой» (англ. light blue). «Исследовав десятки европейских языков, с одной стороны, и языков целого ряда примитивных культур, с другой, Берлин и Кей открыли “всеобщий” эволюционный закон, который выглядит так. Существует семь ступеней (стадий) развития цветовой терминологии, отражающих строгую последовательность появления в лексике языка каждого слова. Более поздняя стадия как бы включает в себя все предыдущие» [Василевич и др., 2005. С. 17].

«На самой “низшей” стадии, стадии I, во всех языках есть только два основных цветообозначения, и это во всех случаях оказываются слова, означающие белый и черный цвета. На стадии II к этим двум прилагательным цвета добавляется третье – и это всегда слово, обозначающее красный цвет. На стадиях III–IV добавляется каждый раз по одному слову из двух – “зелёный”, “жёлтый” (порядок может меняться). На стадии V появляется слово “синий”, на VI – “коричневый”, а высшая стадия VII характеризуется появлением сразу четырех прилагательных цветов – “розовый”, “оранжевый”, “фиолетовый” и “серый» [Там же] (рис. 2).

Концепция Берлина и Кея, сформулированная в 1969 г., не раз пересматривалась и уточнялась [Kay, McDaniel, 1978; Kay, Regier, 2006]. Есть языки, в которых последовательность появления слов, называющих цвета, не вполне соответствует закономерностям, зафиксированным у Берлина и Кея (например, китайский).

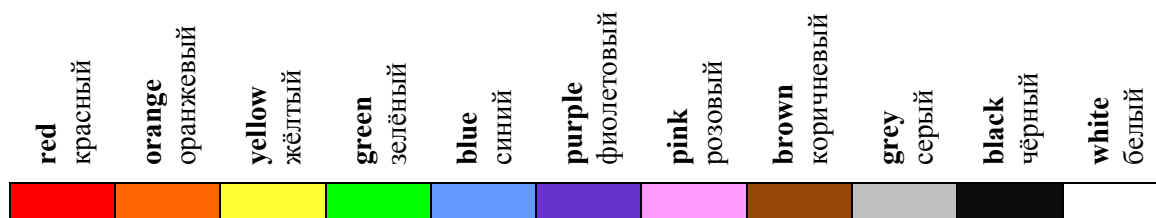


Рис. 1. Базовые цветовые категории по Берлину и Кею

Fig. 1. Basic color categories according to Berlin and Kay

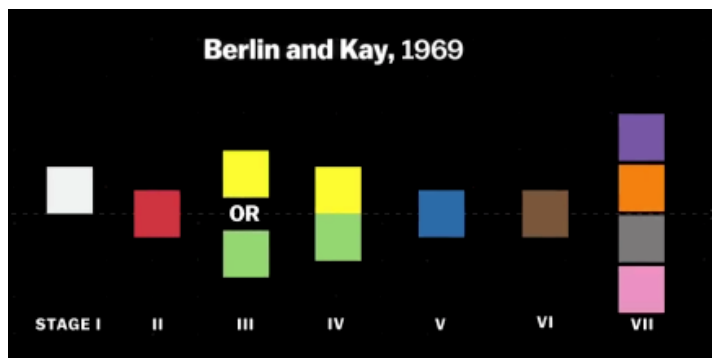


Рис. 2. Графическое изображение эволюции обозначения цветов в лексике английского языка по Берлину и Кею

Fig. 2. Graphic chart of evolution of color designations according to the theory of Berlin and Kay

В 1999 г. Пол Кей и Луиза Меффи опубликовали статью «Появление цвета и возникновение и эволюция лексикона основных цветов» [Kay, Maffi, 1999], в которой описывается серия поправок к теории на основе данных, полученных при реализации проекта «Мировой обзор цвета (World Color Survey – WCS)». Этот проект был предпринят с целью подтвердить, опровергнуть или, что вероятнее всего, модифицировать два основных положения теории Берлина – Кея: (1) наличие универсальных кросс-лингвистических закономерностей именования цветов и (2) фиксированный порядок возникновения системы цветовых терминов. В ходе реализации проекта в базу данных были собраны обозначения цветов из 110 языков. Кроме данных WCS, известна статья Стивена Левинсона, изучавшего вопрос на материале языка Yéí Dnye (Новая Гвинея, Папуа) [Levinson, 2000]. В этой работе отмечены некоторые отличия систем именования цветов от большинства закономерностей, выявленных WCS.

Мы высказали предположение о том, что между последовательностью стадий эволюции языка (по схеме Берлина – Кея, 1969), на которых появляются прилагательные цвета, и частотой их употребления в современном языке имеется корреляция.

Частота употребления базовых прилагательных цвета в корпусах НКРЯ

В данной части исследования мы хотим проверить предположение о том, что порядок появления 11 базовых «цветных» прилагательных (категорий по Берлину и Кею) в диахронии находит отражение в частоте их употребления в синхронном срезе языка, в данном случае русского.

Для подсчета частот мы выбрали корпуса НКРЯ, в первую очередь, основной, а также поэтический, газетный и устный (табл. 1). Дополнительно мы приводим ipm отобранных прилагательных по Новому частотному словарю русской лексики (НЧСРЛ) О. Н. Ляшевской и С. А. Шарова, созданному на базе основного корпуса НКРЯ 2007 г. [Ляшевская, Шаров, 2009].

Рассмотрим порядок снижения частотности прилагательных в корпусах НКРЯ (табл. 2). Как видим, из 11 прилагательных в четырех корпусах (включая НЧСРЛ) самым частотным оказывается «белый», в одном – «чёрный». Во всех пяти корпусах три прилагательных «белый», «чёрный» и «красный» составляют тройку самых частотных. Наименее частотными оказываются прилагательные «розовый», «коричневый», «оранжевый», «фиолетовый». Данные не позволяют говорить о полной корреляции убывания частоты употребления базовых «цветных» прилагательных в корпусах русского языка с эволюционной схемой Берлина и Кея, но, тем не менее, значительная степень корреляция, на наш взгляд, имеется.

Таблица 1

Частотность базовых прилагательных цвета в корпусах НКРЯ

Table 1

Frequencies of the basic adjectives of color in the corpora of the Russian National Corpus (RNC)

Лемма	ОК		ПК		ГК		УК		НЧСРЛ ipm
	с/у	ipm	с/у	ipm	с/у	ipm	с/у	ipm	
Белый	111 109	392,01	10 730	978,37	54 079	236,65	2 115	174,60	339,6
Чёрный	106 635	376,23	9 556	871,33	42 210	184,71	2 286	188,72	337,5
Красный	86 762	306,11	5 582	508,97	47 301	206,99	1 958	161,64	240,5
Синий	29 371	103,63	4 951	451,44	8 722	38,17	450	37,15	82,8
Зелёный	37 891	133,69	4 871	444,14	17 006	74,42	854	70,50	124,4
Голубой	23 412	100,30	4 221	384,88	13 967	61,11	441	36,41	71,5
Серый	38 297	135,12	2 692	245,46	7 753	33,93	453	37,40	96,1
Жёлтый	26 833	94,67	2 423	220,93	8 445	36,95	388	32,03	78,1
Розовый	16 436	57,99	1 883	171,69	3 938	17,23	216	17,83	49,3
Оранжевый	3 944	13,92	260	23,71	6 825	29,87	67	5,53	15,60
Коричневый	6 960	24,56	181	16,50	1 491	6,52	110	9,08	25,5
Фиолетовый	3 325	11,73	125	11,40	727	3,18	57	4,71	12,4
Среднее значение ipm		145,83		360,74		77,49		64,63	122,7

Условные обозначения (Legend):

ПК – поэтический корпус (poetic corpus)

ОК – основной корпус (basic corpus)

ГК – газетный корпус (newspaper corpus)

УК – устный корпус (oral corpus)

НЧСРЛ – Новый частотный словарь русской лексики (Frequency dictionary of the modern Russian language (FDMRL))

с/у – число словоупотреблений (число вхождений при поиске в корпусе) (number of matches (results of search in the corpus))

ipm – instances per million (относительная частота, умноженная на миллион)

Таблица 2

Порядок убывания частоты прилагательных цвета в четырех корпусах НКРЯ и НЧСРЛ
в сопоставлении с эволюционной схемой Берлина и Кея

Table 2

Order of decreasing frequencies of color adjectives in four corpora of the RNC
and in the Frequency dictionary of the modern Russian language (FDMRL) as compared with the evolutionary scheme of Berlin-Kay

Корпус (Corpus)	Ранг (Rank)										
	1	2	3	4	5	6	7	8	9	10	11
Поэтический (poetry)	белый	чёрный	красный	синий	зелёный	серый	жёлтый	розовый	оранжевый	коричневый	фиолетовый
Основной (basic)	белый	чёрный	красный	серый	зелёный	синий	жёлтый	розовый	коричневый	оранжевый	фиолетовый
Газетный (newspaper)	белый	красный	чёрный	зелёный	синий	жёлтый	серый	оранжевый	розовый	коричневый	фиолетовый
Устный (oral)	чёрный	белый	красный	зелёный	серый	синий	жёлтый	розовый	коричневый	оранжевый	фиолетовый
НЧСРЛ (FDMRL)	белый	чёрный	красный	зелёный	серый	синий	жёлтый	розовый	коричневый	оранжевый	фиолетовый
	Стадия (Phase)										
	I	II		III-IV		V	VI	VII			
Берлин – Кей (Berlin – Kay)	белый	чёрный	красный	зелёный	жёлтый	синий	коричневый	серый	розовый	оранжевый	фиолетовый

Три самых частотных прилагательных соответствуют первой и второй стадиям схемы. Правда, полное соответствие наблюдается не во всех корпусах. В одном случае – газетный корпус – прилагательное «красный» оказывается вторым по частотности, а прилагательное «чёрный» третьим. Если рассматривать седьмую стадию, то частотность трех прилагательных из четырех коррелирует со схемой. Полное отсутствие корреляции со схемой наблюдается в частотном поведении прилагательных «серый», «жёлтый» и «коричневый». Частотность первого из них занимает 4–6 ранги в разных корпусах, что, по нашему предположению, соответствует III–V стадиям эволюционной схемы. Напротив, прилагательное «жёлтый» «отстает» от своей позиции в эволюционной схеме. Его частотность в корпусах соответствует V–VI стадиям, а по Берлину – Кею прилагательное «жёлтый» появляется на III–IV стадиях. Частотность прилагательного «коричневый» соответствует, по нашим результатам, VII стадии, а по схеме – VI. Частотность прилагательного «синий» в трех случаях коррелирует со схемой – в основном, устном корпусах и в НЧСРЛ.

Таким образом, мы можем утверждать, что между частотностью прилагательных цвета в разных корпусах и эволюционной схемой Берлина – Кея наблюдается частичная корреляция. В естественном языке, особенно в семантике, вряд ли можно найти явление, которое описывается единой универсальной формулой. Обратим внимание, что порядок убывания частотности прилагательных цвета в разных корпусах хоть и отличается, но отличия в целом не слишком велики, что говорит о наличии закономерностей в частотном распределении прилагательных этой группы.

Корреляция между эволюционной теорией обозначения цвета Берлина и Кея и частотой встречаемости прилагательных цвета в описаниях коллекций Государственного Эрмитажа

Мы предположили, что в описаниях произведений искусства прилагательные цвета будут встречаться чаще в своем основном значении. Поискковая система Эрмитажа не является лингвистическим корпусом. Мы не можем сравнивать относительную частоту встречаемости лексических единиц (irm), поскольку у нас нет данных об общем числе словоупотреблений в системе. Однако можно и интересно сопоставить порядок убывания абсолютного числа найденных описаний по запросам на «цветные» прилагательные. Приведем список результатов поиска по запросу на прилагательные цвета в БД Государственного Эрмитажа (табл. 3). Материал отсортирован по убыванию числа найденных описаний.

Обратим внимание, что малочастотное в корпусах НКРЯ прилагательное «коричневый» в данном случае оказывается третьим по частотности, а частота прилагательного «белый» (самого частотного в корпусах НКРЯ) намного ниже. Как и в приведенных выше данных, можно, пожалуй, говорить о некоторой корреляции со схемой Берлина и Кея. При этом корреляция, против нашего ожидания, выражена слабее, чем в текстах из НКРЯ. Возможно, это особенность коллекций Эрмитажа, возможно, цветовая гамма «мира вещей» чем-то принципиально отличается от цветовой картины «мира языка».

Частотное поведение прилагательных цвета в поэтических текстах

Перейдем к описанию поэтических текстов. Сравним частотное поведение прилагательных цвета в поэтических текстах с текстами других типов.

Уже при анализе данных табл. 1 обращает на себя внимание тот факт, что в целом irm прилагательных цвета в поэтическом корпусе превышает irm прилагательных в остальных трех корпусах, что хорошо видно по средним значениям irm. Лишь в двух случаях (прилагательные «коричневый» и «фиолетовый») значения irm в основном корпусе незначительно превышают irm в поэтическом корпусе, и в еще одном случае irm прилагательного «оранже-

вый» в газетном корпусе превышает частотность этого прилагательного в поэтическом корпусе.

Повышенная частотность прилагательных цвета в поэтическом корпусе особенно наглядно видна, если взять в рассмотрение все «цветные» прилагательные во всех корпусах (табл. 4, условные обозначения см. в табл. 1). Поиск прилагательных цвета производился с использованием семантического тега «цвет».

Таблица 3

Результаты поиска по прилагательным цвета
в поисковой системе Эрмитажа

Table 3

Representation of color adjectives
in the Information System of the State Hermitage Museum

№	Лемма	Число описаний, содержащих данное прилагательное
1	Чёрный	1 186
2	Красный	1 087
3	Коричневый	966
4	Зелёный	917
5	Белый	834
6	Серый	819
7	Жёлтый	367
8	Голубой	326
9	Синий	320
10	Розовый	219
11	Фиолетовый	43

Таблица 4

Частотность прилагательных, имеющих тег семантической разметки «цвет»,
в четырех корпусах НКРЯ

Table 4

Frequency of adjectives supplied with the tag “color”
in four corpora of the RNC

ПК		ОК		ГК		УК	
с/у	ipm	с/у	ipm	с/у	ipm	с/у	ipm
85 286	7776,48	77 1638	2722,48	280 630	1228,02	12 682	1046,93

Анализ данных табл. 4 подтверждает сделанные ранее наблюдения, что частотность прилагательных цвета в текстах поэтического корпуса выражена выше, чем в остальных трех корпусах.

Методика исследования

Для того чтобы получить статистические данные, характеризующие употребление прилагательных цвета в русском поэтическом языке и в языке отдельных авторов, была выполнена следующая работа.

1. В поэтическом корпусе НКРЯ на первом шаге созданы авторские подкорпуса из текстов 15 авторов. Были выбраны наиболее известные авторы XIX–XX вв. из числа представленных в НКРЯ. В созданных корпусах производится поиск прилагательных цвета по семантическому тегу «цвет». В конце каждой страницы результатов поиска корпусный менеджер НКРЯ выводит таблицы частоты встречаемости заданных в запросе лексических единиц в подмножестве документов данной страницы результатов, отдельно для словоформ и лемм. Нас интересует таблица частоты лемм.

2. Таблицы словоформ и лемм указывают частоты лексем только по данной странице результатов, а таких страниц может оказаться немало. Поэтому каждая таблица лемм для каждой очередной страницы результатов поиска переносится в книгу MS Excel, где данные суммируются средствами Excel («Сводная таблица»), т. е. для каждой леммы подсчитывается суммарное количество словоупотреблений.

3. Средствами Excel для каждой леммы данного авторского подкорпуса подсчитываются дополнительные параметры (величина *ipm* по отношению к объему всего авторского подкорпуса и доля данного прилагательного (леммы) среди всех прилагательных цвета в текстах у данного автора) (рис. 3).

4. Наконец, таблицы по отдельным авторам сливаются в общую таблицу, в которой выполняются сортировки и анализ данных по различным критериям.

5. Также для вычисления показателей использования прилагательных цвета каждым автором запоминаются (вычисляются) объем каждого авторского корпуса, число цветных прилагательных, число всех прилагательных в этом корпусе.

1	Лемма	автор	Число с/у	IPM	%
2	белый	Блок	225	1982,5	13,69
3	темный	Блок	193	1700,55	11,74
4	черный	Блок	182	1603,62	11,07
5	светлый	Блок	159	1400,97	9,67
6	красный	Блок	131	1154,26	7,97
7	синий	Блок	118	1039,71	7,18
8	бледный	Блок	103	907,54	6,27
9	голубой	Блок	97	854,68	5,9
10	зеленый	Блок	73	643,21	4,44
11	серый	Блок	52	458,18	3,16
12	алый	Блок	41	361,26	2,49
13	розовый	Блок	41	361,26	2,49
14	лазурный	Блок	38	334,82	2,31
15	желтый	Блок	33	290,77	2,01
16	сумрачный	Блок	26	229,09	1,58
17	пестрый	Блок	17	149,79	1,03
18	золотистый	Блок	16	140,98	0,97

Рис. 3. Фрагмент таблицы частотности прилагательных цвета в текстах А. А. Блока

Fig. 3. Fragment of the frequency table of color adjectives in the texts by Alexander Block

В поэтическом корпусе НКРЯ отсутствуют тексты многих известных русских поэтов второй половины XX в., таких как Евтушенко, Вознесенский, Ахмадулина, Высоцкий, Окуджава, Матвеева и др. В связи с этим на базе различных массивов текстов, взятых из Интернета (Библиотека Максима Мошкова, персональные сайты поэтов) средствами менеджера Sketch Engine были созданы авторские корпуса текстов других русских поэтов (еще 21 автор). В этой системе нет семантической разметки, поэтому в этих новых корпусах поиск производился по сводному списку прилагательным цвета, полученному на основе НКРЯ.

Далее по корпусу каждого автора выполняются операции 3–5, описанные выше.

Частотное поведение прилагательных цвета в поэтических текстах отдельных авторов

В конечном счете, все таблицы для 36 авторов были сведены в единую таблицу Excel общим объемом строк 1571. Таблица 5 представляет собой фрагмент сводной таблицы для 50 лемм с наибольшим *ipm* в корпусах отдельных авторов (сортировка по *ipm*).

Таблица 5

Фрагмент сводной таблицы частотности прилагательных,
обозначающих цвет, в текстах 36 авторов (часть 1)

Table 5

Fragment of the summary frequency table of color adjectives
in the texts by 36 authors (part 1)

Ранг и лемма	Автор	Число с/у	<i>ipm</i>	Доля, %
1. Белый	Хлебников	223	4041,10	22,57
2. Черный	Хлебников	149	2700,11	15,08
3. Белый	Ахматова	122	2159,14	18,63
4. Черный	Набоков	130	2079,87	13,50
5. Черный	Ахматова	117	2070,65	17,86
6. Синий	Хлебников	111	2011,49	11,23
7. Белый	Блок	225	1982,50	13,69
8. Белый	Ахмадулина	103	1842,64	28,77
9. Белый	Вознесенский	52	1822,45	25,00
10. Синий	Есенин	105	1789,61	14,73
11. Белый	Есенин	101	1721,43	14,17
12. Темный	Блок	193	1700,55	11,74
13. Белый	Тарковский	63	1696,01	18,81
14. Черный	Мандельштам	102	1689,05	16,59
15. Светлый	Бальмонт	179	1655,90	13,81
16. Белый	Набоков	102	1631,90	10,59
17. Черный	Блок	182	1603,62	11,07
18. Темный	Хлебников	88	1594,69	8,91
19. Белый	Бальмонт	166	1535,64	12,81
20. Белый	Кривулин	44	1513,48	18,64
21. Белый	Заболоцкий	31	1464,27	21,09

Окончание табл. 5

Ранг и лемма	Автор	Число с/у	ipm	Доля, %
22. Черный	Тарковский	54	1453,72	16,12
23. Черный	Вознесенский	41	1436,93	19,71
24. Красный	Цветаева	245	1435,59	15,88
25. Белый	Цветаева	240	1406,30	15,55
26. Светлый	Блок	159	1400,97	9,67
27. Белый	Рубцов	35	1353,70	15,56
28. Черный	Есенин	78	1329,42	10,94
29. Светлый	Тютчев	45	1325,24	24,46
30. Темный	Рубцов	34	1315,03	15,11
31. Черный	Цветаева	222	1300,82	14,39
32. Черный	Кривулин	36	1238,3	15,25
33. Зеленый	Хлебников	68	1232,26	6,88
34. Златой	Державин	115	1229,91	18,34
35. Голубой	Хлебников	67	1214,14	6,78
36. Голубой	Вертинский	16	1196,53	16,00
37. Белый	Аронзон	34	1179,33	24,29
38. Красный	Блок	131	1154,26	7,97
39. Красный	Маяковский	156	1153,09	26,17
40. Бледный	Бальмонт	122	1128,61	9,41
41. Голубой	Есенин	65	1107,85	9,12
42. Черный	Пастернак	86	1104,43	21,39
43. Зеленый	Тарковский	41	1103,75	12,24
44. Белый	Пастернак	81	1040,22	20,15
45. Синий	Блок	118	1039,71	7,18
46. Темный	Мандельштам	62	1026,68	10,08
47. Черный	Симонов	84	1020,97	21,65
48. Светлый	Рубцов	26	1005,61	11,56
49. Темный	Бальмонт	107	989,84	8,26
50. Красный	Есенин	58	988,55	8,13

Из 25 наиболее частотных словоупотреблений (ранги 1–25) 12 прилагательных «белый», 7 – «черный», 2 – «синий», 1 – «красный». Еще три словоупотребления: прилагательные «темный» и «светлый», являются в нашей терминологии «квазицветами» (т. е. в семантической разметке НКРЯ имеют тег «цвет» и потому входят в массив обнаруженных слов). Здесь можно заметить некоторую корреляцию между частотностью прилагательных цвета и стадиями эволюции цветообозначений. В данном частотном ранге преобладают слова, обозначающие цвета I эволюционной стадии. На 24-м месте появляется «красный» – цвет, отнесенный Берлином и Кеем к стадии II. Несколько нарушает корреляцию прилагательное «синий» у Есенина и Хлебникова, но это, по-видимому, особенности авторских идиолектов.

В следующем фрагменте (ранги 26–50) преобладает прилагательное «чёрный» (5 позиций), появляются цвета, отсутствовавшие в первом фрагменте таблицы: «зелёный» (2 раза), «голубой» (3 раза), появилось аналоговое прилагательное «золотой».

Сформируем табл. 6, показывающую порядок появления основных цветовых категорий в сводной таблице (1 571 строка).

Таблица 6

Ранг первого появления основных цветовых категорий
в сводной таблице

Table 6

Frequency ranking of the first appearance of basic color categories
in the summary table

Лемма	Автор	Ранг первого появления в списке	ipm
Белый	Хлебников	1	4041,1
Черный	Хлебников	2	2700,11
Синий	Хлебников	6	2011,49
Красный	Цветаева	23	1435,59
Зеленый	Хлебников	33	1232,26
Голубой	Хлебников	35	1214,14
Желтый	Ахматова	82	725,61
Серый	Хлебников	101	634,25
Розовый	Набоков	133	511,97
Оранжевый	Вознесенский	252	280,38
Коричневый	Мандельштам	582	66,24
Фиолетовый	Набоков	593	64,00

Можно говорить о некоторой (неполной) корреляции между рангами прилагательных в табл. 6 и их принадлежностью к эволюционным стадиям. Обращает внимание, что из 12 прилагательных 6 впервые встречаются в текстах В. Хлебникова. Это, по-видимому, говорит о каких-то существенных особенностях текстов этого автора. Однако анализ текстов отдельных авторов не входит в нашу задачу.

Классификация имен прилагательных цвета

Как мы уже говорили, таблицы 36 авторов были слиты в единую таблицу (1 571 строка). Для 180 лемм, представленных в этой общей таблице, было подсчитано число авторов, у которых каждая из лемм встретилась (табл. 7).

Для обозначения цвета естественный язык использует разные части речи. В данной статье рассматриваются только прилагательные. Но и для прилагательных цвета существуют различные классификации [Герасимов, 1969; 1978; Василевич и др., 2005]. Рискнем предложить еще одну. Как часто бывает, в рамках одной классификации присутствуют разные основания выделения классов. Этот недостаток свойственен и нашей. Но главная наша задача – дать статистическую картину широкой или ограниченной распространенности различных прилагательных и их типов в текстах разных поэтов и в поэтическом языке в целом.

Таблица 7
Наиболее часто встречающиеся
прилагательные цвета

Table 7
Most frequent color adjectives

Лемма	Число авторов
1. Белый	36
2. Голубой	36
3. Желтый	36
4. Зеленый	36
5. Красный	36
6. Светлый	36
7. Синий	36
8. Темный	36
9. Черный	36
10. Бледный	35
11. Серый	34
12. Розовый	33
13. Рыжий	31
14. Цветной	31
15. Багровый	30
16. Пестрый	30
17. Алый	29
18. Мутный	29
19. Белоснежный	28
20. Золотистый	26
21. Лиловый	24
22. Сизый	24
23. Сумрачный	24
24. Серебристый	23
25. Багряный	21

Вот предлагаемая нами классификация.

1. Прилагательные, означающие собственно цвета
2. Прилагательные, означающие квазицвета
3. Прилагательные, означающие специфические цвета определенных объектов
4. Прилагательные цвета-дериваты
 - 4.1. Диминутивы
 - 4.2. Прилагательные подобия
 - 4.3. Составные прилагательные цвета
5. Аналоговые цвета

Первая группа – собственно цвета. Таких слов мы обнаружили 29 из 180 (15,7 %) (табл. 8).

Таблица 8

Прилагательные, означающие собственно цвет

Table 8

Adjectives, meaning color per se

Лемма	Число авторов	Лемма	Число авторов
1. Белый	36	16. Бурый	18
2. Голубой	36	17. Оранжевый	18
3. Желтый	36	18. Пурпурный	16
4. Зеленый	36	19. Коричневый	13
5. Красный	36	20. Лазоревый	13
6. Синий	36	21. Фиолетовый	12
7. Черный	36	22. Пунцовый	10
8. Серый	34	23. Пурпуровый	10
9. Розовый	33	24. Сивый	10
10. Багровый	30	25. Палевый	9
11. Алый	29	26. Бордовый	5
12. Лиловый	24	27. Червлёный	5
13. Сизый	24	28. Бежевый	3
14. Багряный	21	29. Бордовый	3
15. Лазурный	20	Среднее число авторов	21,1

В эту группу вошли все 12 основных цветов. В пределах одного частотного ранга слова отсортированы по алфавиту. И в этом случае можно говорить о некоторой корреляции между степенями эволюции по Берлину и Кею.

Среди прилагательных этой группы встречаются синонимы. Вряд ли можно сделать уверенное различие между значениями прилагательных «красный», «алый», «пунцовый», «бордовый» или «фиолетовый», «лиловый», а также, например, «бурый» и «коричневый». Подтвердим этот тезис с помощью поисковой системы Государственного Эрмитажа – вот некоторые образцы синонимичных цветов (рис. 4–6).

Представленные образцы демонстрируют похожесть и различия между цветами, обозначаемыми одним прилагательным, и особенности каждого оттенка, которые, кстати говоря, трудно поддаются вербализации, особенно с точки зрения рядового носителя языка. Так, цвета предметов на изображениях, полученных по запросу «красный», субъективно воспринимаются как различные, что, возможно, обусловлено различием фактуры материалов. С другой стороны, цвета предметов на изображениях, полученных по запросу «алый» и «бордовый», субъективно воспринимаются скорее как сходные, несмотря на различную фактуру материала.

Нет сомнения, что авторы описаний коллекций Эрмитажа, являясь профессиональными искусствоведами, дают предельно точные наименования цветов. В этом случае, видимо, нужно говорить об особенностях языков для специальных целей. Кроме того, прилагательное «красный» обозначает не просто цвет, а широкую цветовую категорию, содержащую большее число значений – оттенков красного цвета. А категории «алый» и «бордовый» значительно уже, индивидуальнее.



Рис. 4. Поисковый термин «красный» (результат поиска – 1 087 описаний):

a – фрагмент: Эннер, Жан-Жак (1829–1905), Этюд женщины в **красном**, Франция, начало 1890-х гг., холст, масло, 55 × 38 см; *б* – обломок фриза из **красного** мрамора с гравировкой, Урарту, VIII–VII вв. до н. э., мрамор, 6 × 2,2 см; *в* – ткань с крупным стилизованным растительным узором по **красному** фону, Испания (?), середина XVII в., шелк, серебряная позолоченная нить, 55 × 53 см

Fig. 4. Search term “red”:

a – fragment: Henner Jean-Jacques (1829–1905), Study of woman in **red** dress, France, beginning of 1890-s; oil, canvas 55 × 38 cm; *б* – piece of frieze made of **red** marble with engraving; Urartu; 9th – 7th century B.C.; marble; 6 × 2,2 cm; *в* – textile with large-sized floral pattern against **red** background; Spain (?), 17th century; silk; silver gilded thread; 55 × 53 cm



Рис. 5. Поисковый термин «алый» (результат поиска 46 описаний):

a – фрагмент: Икона – Святой Николай, место создания Россия, Новгородская школа, XIII–XIV вв., дерево, темпера, 107 × 82 × 3 см (из аннотации: «**алый (киноварный)** фон (воплощение негаснущего света вечности), характерный для новгородской школы иконописи...»); *б* – мундирное платье Екатерины II, по форме армейской пехоты, **шелк алый**, рукава и спина зеленого шелка, подкладка шелковая (на лифе – белая, на юбке – алая), галуны и пуговицы золоченые (сзади и спереди разрезы для езды верхом по-мужски), 1772 г., шелк

Fig. 5. Search term “scarlet”:

a – fragment: Icon Saint Nicolas; Russia; the Novgorodian school; 13th – 14th century; wood; tempera; 107 × 82 × 3 cm (from annotation: “**scarlet (cinnabar)** background (symbol of inextinguishable light of eternity) typical of the Novgorodian school”); *б* – uniform dress of Catharina II, to the uniform of army infantry, scarlet silk, sleeves and back of green silk, backing of silk (white on the bodice, scarlet on the skirt) gallons and buttons gilded, 1772; silk



а



б

Рис. 6. Поисковый термин «бордовый» (результат поиска 8 описаний):

а – фрагмент: Винтерхальтер, Франц Ксавер. 1806–1873, Портрет С. П. Нарышкиной, Франция, 1858 г., холст, масло 150 × 114 см (из аннотации: «...изящный жест холеных рук, смелое сочетание темно-**бордовых** и лимонно-желтых тонов одежды придают облику модели претенциозную элегантность»); б – передвижная отопительная печь, Франция, конец XIX в., чугун, эмаль, металл, слюда литые, ковка, чеканка 72 × 68 × 33 см (Печь, покрытая эмалью бордового цвета, по форме и рельефному декору относится к эпохе модерна)

Fig. 6. Search term “vinous”:

а – fragment: Winterhalter Franz Xaver (1806–1873), Portrait of S. P. Naryshkina; France 1858, oil, canvas (from annotation: “daring combination of dark **vinous** and lemon yellow colors of clothes enhance to the image pretentious elegance...”); б – movable heating stove; France; end of 19th century, cast, iron, metal, mica, molding, forging, embossed work; 72 × 68 × 33 cm (The stove covered with enamel of **vinous** color...)

Вторую группу слов мы условно назвали «квазицвета» (19 лемм, 10,6 %) (табл. 9).

Таблица 9

Прилагательные «квазицвета»

Table 9

“Quasi colors” adjectives

Лемма	Число авторов
1. Светлый	36
2. Тёмный	36
3. Бледный	35
4. Цветной	31
5. Пёстрый	30
6. Мутный	29
7. Сумрачный	24
8. Разноцветный	21
9. Бесцветный	20
10. Клетчатый	10
11. Самоцветный	10

Окончание табл. 9

Лемма	Число авторов
12. Пятнистый	9
13. Цветастый	8
14. Линялый	6
15. Трёхцветный	3
16. Двухцветный	1
17. Крапчатый	1
18. Одноцветный	1
19. Потускнелый	1
Среднее число авторов	16,4

В число прилагательных, обозначающих цвета (т. е. имеющих тег «цвет» в НКРЯ), попадают слова «тёмный», «светлый», «бледный», «пёстрый», «сумрачный», «цветной», которые, собственно, означают не цвет, а его интенсивность, оттенок (тёмный, светлый, бледный), просто наличие цвета (цветной) – как оппозиция черному и белому, множество цветов (пестрый) или даже настроение (сумрачный), которое могут вызвать цвет или комбинация цветов. Эти прилагательные описывают различные визуальные характеристики объекта – кроме, пожалуй, его формы.

Часть этих слов не обозначает, а характеризует цвет. Некоторые из них образуют сложные прилагательные типа «бледно-зелёный», «светло-серый», «тёмно-красный» и т. д. (см. далее, табл. 15).

Как видно из табл. 9, представленность в текстах некоторых из этих прилагательных достаточно высока. Особенно это касается прилагательных «светлый» и «тёмный». Некоторые авторы утверждают, что именно смена светлого и темного времени суток обусловила появление цветообозначений «белый» и «чёрный» на самой ранней стадии эволюции (см. рис. 2).

Сравним результаты поиска этого вида прилагательных в разных корпусах НКРЯ (табл. 10).

По результатам анализа отметим следующее.

1. Средний *ipm* названий квазицветов ниже, чем названий цветов (см. табл. 2):

- по поэтическому корпусу в 3,4 раза;
- по основному корпусу в 5,1 раза;
- по газетному корпусу в 11,4 раза;
- по устному корпусу 8,4 раза.

2. Как и в случае с основными цветами среднее *ipm* квазицветов в поэтическом корпусе значительно выше, чем в трех других корпусах.

Третью группу составляют прилагательные, которые относятся к определенному, чаще всего одушевленному, предмету. Это может быть цвет волос или глаз, масть или окрас шерсти животного и т. п. (28 лемм, 15,5 %) (табл. 11, 12).

Тенденция соотношений *ipm* у этого типа прилагательных примерно та же, что у прилагательных собственно цвета и прилагательных «квазицветов», т. е. среднее *ipm* в поэтическом корпусе значительно выше, чем в остальных трех корпусах. При этом среднее *ipm* этого типа прилагательных во всех корпусах ниже, чем для прилагательных собственно цвета и квазицветов.

Таблица 10

Сравнение частотности прилагательных «квазицвета» в текстах четырех корпусов НКРЯ

Table 10

Comparison of frequency of the “quasi colors” adjectives in texts of four corpora the RNC

Лемма	ПК		ОК		ГК		УК	
	с/у	ipm	с/у	ipm	с/у	ipm	с/у	ipm
Тёмный	7310	666,53	60395	213,08	8899	38,94	820	67,69
Светлый	6631	604,62	33797	119,24	8097	35,43	515	42,51
Бледный	3910	356,52	20243	71,42	1482	6,49	90	7,43
Пёстрый	1100	100,30	7581	26,75	913	4,00	34	2,81
Сумрачный	942	85,89	2972	10,49	180	0,79	11	0,91
Мутный	835	76,14	7116	25,11	964	4,22	41	3,38
Цветной	585	53,34	9060	31,97	5918	25,90	166	13,70
Разноцветный	276	25,17	5272	18,60	1508	6,60	24	1,98
Бесцветный	166	15,14	2580	9,10	250	1,09	17	1,40
Самоцветный	86	7,84	289	1,02	14	0,06	3	0,25
Пятнистый	65	5,93	964	3,40	249	1,09	8	0,66
Клетчатый	46	4,19	1881	6,64	334	1,46	27	2,23
Линялый	41	3,74	306	1,08	41	0,18	2	0,17
Трёхцветный	41	3,74	520	1,83	268	1,17	9	0,74
Цветастый	14	1,28	623	2,20	260	1,14	4	0,33
Одноцветный	14	1,28	256	0,90	39	0,17	2	0,17
Потускнелый	5	0,46	12	0,04	0	0,00	0	0,00
Крапчатый	4	0,36	122	0,43	9	0,04	2	0,17
Двухцветный	3	0,27	124	0,44	88	0,39	4	0,33
Среднее значение ipm		105,9		28,62		6,8		7,73
Среднее значение для прилагательного собственно цвета (см. табл. 1)		360,74		145,83		77,48		64,63

Таблица 11

Прилагательные,
относящиеся к определенному предмету

Table 11

Object-related adjectives

Лемма	Число авторов
1. Рыжий	31
2. Румяный	21
3. Русый	21
4. Гнедой	15
5. Карий	14
6. Пегий	10
7. Вороной	9
8. Поседельй	9
9. Рыжеватый	7
10. Седенький	7
11. Сизокрылый	7
12. Буланный	6
13. Порыжелый	4
14. Рыжеволосый	4
15. Рыженький	4
16. Рябоватый	3
17. Саврасый	3
18. Чалый	3
19. Черноволосый	3
20. Черномазый	3
21. Седоватый	2
22. Соловый	2
23. Каурый	1
24. Муругий	1
25. Мухоргый	1
26. Рыжебородый	1
27. Рябенький	1
28. Чубарый	1
Среднее число авторов	6,9

Frequency of the object-related adjectives in four corpora of the RNC

Лемма	ПК		ОК		ГК		УК	
	с/у	ipm	с/у	ipm	с/у	ipm	с/у	ipm
Румяный	992	90,45	4159	14,67	442	1,93	23	1,90
Рыжий	931	84,89	14326	50,54	1780	7,79	234	19,32
Русый	311	28,36	2919	10,30	1326	5,80	29	2,39
Карий	136	12,40	2164	7,63	720	3,15	28	2,31
Вороной	117	10,67	1467	5,18	246	1,08	32	2,64
Пегий	113	10,30	772	2,72	22	0,10	2	0,17
Рыжебородый	95	8,66	450	1,59	11	0,05	1	0,08
Гнедой	70	6,38	907	3,20	95	0,42	18	1,49
Поседельй	60	5,47	459	1,62	80	0,35	7	0,58
Буланый	39	3,56	221	0,78	4	0,02	1	0,08
Соловый	38	3,46	197	0,70	15	0,07	0	0,00
Сизокрылый	33	3,01	50	0,18	10	0,04	1	0,08
Рыжеватый	31	2,83	1242	4,38	59	0,26	4	0,33
Седенький	24	2,19	477	1,68	25	0,11	3	0,25
Рыженький	22	2,01	671	2,37	71	0,31	21	1,73
Чалый	21	1,91	304	1,07	221	0,97	0	0,00
Черноволосый	20	1,82	1086	3,83	130	0,57	6	0,50
Саврасый	18	1,64	121	0,43	2	0,01	0	0,00
Рыжеволосый	17	1,55	375	1,32	210	0,92	0	0,00
Черномазый	16	1,46	410	1,45	19	0,08	9	0,74
Каурый	13	1,19	132	0,47	12	0,05	2	0,17
Порыжелый	12	1,09	219	0,77	0	0,00	0	0,00
Седоватый	9	0,82	410	1,45	16	0,07	2	0,17
Рябоватый	6	0,55	218	0,77	6	0,03	0	0,00
Муругий	6	0,55	27	0,10	0	0,00	0	0,00
Рябенький	3	0,27	89	0,31	9	0,04	3	0,25
Чубарый	3	0,27	56	0,20	3	0,01	0	0,00
Мухортый	1	0,09	29	0,10	0	0,00	0	0,00
Среднее значение IPM		10,28		4,28		0,86		1,26

Некоторая часть этих слов также не означает собственно цвета, но цветовую неоднородность, наличие пятен («пегий», «рябенький», «чубарый», «чалый»), в других случаях указывается на результат некоторого процесса («поседелый», «порыжелый»). Иногда образуется прилагательное, содержащее одновременно объект и цветовую характеристику, причем и то, и другое присуще одному лицу (рыжебородый, черноволосый).

Значительное место в этой группе занимает масть лошадей или окрас собак. Известно, что масть или окрас – это не столько цвет, сколько сочетание цветов, генетически детерминированный тип распределения пигментов в шерсти животного.

Четвертая группа прилагательных цвета, дериваты (83 леммы, 46 %), может быть разделена на три подгруппы, отражающие некоторые формы прилагательных. Так, в первую подгруппу четвертой группы вошли прилагательные-диминывы (13 лемм, 7,2 %) (табл. 13).

Таблица 13
Прилагательные-диминывы
Adjectives diminutive

Table 13

Лемма	Число авторов
1. Беленький	13
2. Серенький	13
3. Синенький	9
4. Аленький	6
5. Голубенький	5
6. Чёрненький	5
7. Зелёненький	3
8. Жёлтенький	2
9. Красненький	2
10. Бледненький	1
11. Темнёхонький	1
12. Чернёхонький	1
13. Чернёшенький	1
Среднее число авторов	4,8

В их число вошли не только собственно прилагательные цвета, но и прилагательные «темнёхонький» и «бледненький», исходные формы которых «тёмный» и «бледный» мы отнесли к квазицветам. В табл. 12 мы находим три прилагательных, от которых образованы диминывы, – «рыженький», «рябенький» и «седенький». Здесь нам пришлось столкнуться с типичной проблемой любой классификации – один объект имеет два значимых классификационных признака.

Интересно, что в поисковой системе Эрмитажа цветоозначающие прилагательные-диминывы отсутствуют. Это связано с тем, что эта форма придает слову экспрессивность, ненужную в научных описаниях произведений искусства.

Прилагательные второй подгруппы условно назовем прилагательными подобия (10 лемм, 5,5 %) (табл. 14).

Таблица 14
Прилагательные подобия
Adjectives of likeness
Table 14

Лемма	Число авторов
1. Розоватый	13
2. Синеватый	11
3. Зеленоватый	9
4. Красноватый	8
5. Голубоватый	7
6. Лиловатый	6
7. Желтоватый	4
8. Сероватый	4
9. Белесоватый	2
10. Беловатый	1
Среднее число авторов	6,5

В этой форме встретились прилагательные, исходные формы которых мы отнесли к третьей группе (см. табл. 12), – «рябоватый», «рыжеватый» и «седоватый». Обращает внимание, что от этих же исходных слов образованы диминутивы (см. табл. 13).

Третья подгруппа – это сложные прилагательные, имеющие обычно два корня (62 леммы, 35 %) (табл. 15). Они разными способами описывают интенсивность или оттенок цвета.

Число лемм сложных прилагательных значительно превышает число членов других групп и подгрупп. При этом частотность каждой леммы ниже, чем у лемм в остальных группах и подгруппах (среднее число авторов на одну лемму 2,26).

Чаще всего в составные прилагательные этого типа входят комбинации с корнями *тёмно-* (9 лемм), *ярко-* (9 лемм), *чёрно-* (7 лемм), *светло-* (6 лемм), *нежно-* (5 лемм), *бледно-* (4 леммы). Отметим, что прилагательные этой группы присутствуют и в поисковой системе Эрмитажа (табл. 16).

Обращает на себя внимание низкая частотность сложных прилагательных по сравнению с прилагательными основных цветов. Наибольшее число описаний найдено на сочетании с прилагательным «коричневый».

Наконец, пятая группа содержит прилагательные, которые мы называли аналоговыми. Они передают цвета и оттенки через цвета и оттенки хорошо известных предметов. Герасимов [1969] называет их относительными прилагательными. Семантическая разметка НКРЯ плохо учитывает слова такого типа, поэтому при поиске по семантическому признаку в корпусе найдено всего 8 таких прилагательных (5 %) (табл. 17). Однако в текстах НКРЯ, как показывают другие наши исследования, их значительно больше, просто большинству из них не приписан тег «цвет». Следует отметить, что частотность каждого из найденных прилагательных довольно высока.

Таблица 15

Сложные прилагательные цвета

Table 15

Compound adjectives of color

Лемма	Число авторов	Лемма	Число авторов
1. Тёмно-синий	14	32. Красно-синий	1
2. Тёмно-голубой	9	33. Кроваво-красный	1
3. Чёрно-белый	9	34. Медно-зелёный	1
4. Тёмно-зелёный	8	35. Небесно-голубой	1
5. Бледно-голубой	7	36. Нежно-зелёный	1
6. Чёрно-синий	6	37. Нежно-золотой	1
7. Светло-зелёный	5	38. Нежно-розовый	1
8. Ярко-жёлтый	4	39. Потёрто-белый	1
9. Молочно-белый	3	40. Пречёрный-чёрный	1
10. Нежно-белый	3	41. Прозрачно-синий	1
11. Светло-серый	3	42. Розовато-белый	1
12. Светло-синий	3	43. Светло-пегий	1
13. Серо-зелёный	3	44. Светло-сиреневый	1
14. Серо-синий	3	45. Серо-голубой	1
15. Ярко-красный	3	46. Сине-красный	1
16. Бледно-зелёный	2	47. Тёмно-бирюзовый	1
17. Бледно-лиловый	2	48. Тёмно-вишневый	1
18. Медно-красный	2	49. Тёмно-лиловый	1
19. Нежно-голубой	2	50. Чёрно-бархатный	1
20. Светло-голубой	2	51. Чёрно-бурый	1
21. Сине-зелёный	2	52. Чёрно-золотой	1
22. Синий-синий	2	53. Чёрно-красной	1
23. Тёмно-жёлтый	2	54. Чёрно-серебряный	1
24. Тёмно-красный	2	55. Электродово-синий	1
25. Тёмно-серый	2	56. Ярко-белые	1
26. Бледно-розовый	1	57. Ярко-голубой	1
27. Буйно-красный	1	58. Ярко-голубой	1
28. Грязно-белый	1	59. Ярко-зелёный	1
29. Грязно-розовый	1	60. Ярко-золотистый	1
30. Жёлто-зелёный	1	61. Ярко-оранжевый	1
31. Зеленовато-голубой	1	62. Ярко-розовый	1
		Среднее число авторов	2,3

Таблица 16

Результаты поиска сложных слов в поисковой системе Эрмитажа

Table 16

Representation of compound color adjectives
in the Information System of the State Hermitage Museum

Лемма	Число описаний, содержащих данное прилагательное	Лемма	Число описаний, содержащих данное прилагательное
1. Тёмно-коричневый	43	15. Ярко-синий	3
2. Светло-коричневый	42	16. Бледно-зелёный	2
3. Тёмно-синий	33	17. Тёмно-голубой	2
4. Тёмно-зелёный	18	18. Тёмно-розовый	2
5. Светло-зелёный	15	19. Бледно-жёлтый	1
6. Ярко-красный	12	20. Бледно-голубой	1
7. Светло-жёлтый	10	21. Бледно-розовый	1
8. Светло-серый	10	22. Бледно-серый	1
9. Тёмно-красный	9	23. Светло-розовый	1
10. Светло-голубой	8	24. Светло-синий	1
11. Светло-фиолетовый	5	25. Ярко-голубой	1
12. Тёмно-серый	4	26. Ярко-зелёный	1
13. Тёмно-фиолетовый	3	27. Ярко-розовый	1
14. Ярко-жёлтый	3	28. Бледно-зелёный	2

Таблица 17

Аналоговые прилагательные цвета

Table 17

Analog adjectives of color

Лемма	Число авторов
1. Золотистый	26
2. Серебристый	23
3. Златой	21
4. Изумрудный	18
5. Бирюзовый	10
6. Дымчатый	8
7. Агатový	3
8. Аметистový	2
Среднее число авторов	13,9

Выводы

Исследование прилагательных цвета в поэтических текстах 36 авторов показывает следующее.

1. Число прилагательных, означающих собственно цвет, составляет 29 единиц. Эти прилагательные наиболее частотны.

2. Что касается частотного поведения прилагательных цвета вообще, то здесь можно обозначить два аспекта. С одной стороны, частота встречаемости прилагательных в разной степени коррелирует с эволюционной теорией цветов Берлина и Кея, т. е. прилагательные, относящиеся согласно этим авторам к более ранней стадии эволюции, являются более частотными. С другой стороны, эта корреляция не является абсолютной.

3. Степень корреляции частотности прилагательных цвета в разных текстах НКРЯ и стадий эволюционной модели Берлина и Кея примерно одинакова. Корреляция между моделью Берлина и Кея и абсолютной частотой прилагательных цвета в поисковой системе Эрмитажа выражена слабее.

4. Частота встречаемости прилагательных цвета любого типа значительно выше в поэтических текстах, чем в устных речевых произведениях, газетных статьях и т. п.

5. Отмечаются выраженные различия частотности прилагательных, относящихся к разным классам нашей классификации.

6. Частотность прилагательных собственно цвета всегда выше частотности прилагательных других классов нашей классификации.

7. Сравнение частотности прилагательных цвета по корпусам поэтических текстов разных авторов показывает большой разброс значений *ipm*.

8. При расположении лемм прилагательных по убыванию частотности (по числу словоупотреблений / лемм в авторских корпусах либо по числу авторов, в текстах которых встречается данная лемма) видно, что основные категории более частотны, и при этом существует очень большое число сложных цветовых терминов единичного употребления (закон Ципфа подтверждается и здесь).

9. Наблюдается корреляция между использованием прилагательных различными авторами.

10. Анализ использования прилагательных цвета (50 самых частотных лемм) в текстах 36 авторов (см. табл. 5) показывает, что в текстах авторов XX в. наблюдается гораздо более высокая частотность прилагательных цвета. Так, в число первых пятидесяти самых больших значений *ipm* (для пары «лемма – автор») из авторов XIX в. вошли только Державин (золотой) и Тютчев (светлый).

11. Прилагательные основных цветов часто имеют большое число синонимов, различия между которыми трудно поддаются вербализации. Однако на основании сравнения образцов, полученных через поисковую систему Эрмитажа, можно предположить, что различия в восприятии основных цветов более выражены, чем различия в восприятии цветов, обозначаемых синонимами.

Заключение

Лингвистические корпуса и компьютерные технологии дают возможность осуществлять исследования и получать результаты на больших текстовых массивах, что прежде, до появления этих средств, было недостижимо.

Наличие в семантической разметке НКРЯ тега «цвет» делает возможным объемное статистическое исследование русских прилагательных, обозначающих цвета. Однако в ходе исследования мы обнаружили, что в ряде случаев этот признак не приписывается цветообозначающему прилагательному (например, аналоговому).

Кроме того, невозможно автоматически установить прямое или переносное значение прилагательного. Существуют, правда, интересные попытки решить такую задачу [Чекменева 2018], но результаты носят пока предварительный характер. Есть и другие нерешенные проблемы семантической разметки, которые в настоящей публикации мы не затрагиваем. В наши дальнейшие планы входит как раз описание и систематизация этих проблем.

Среднее значение *ipm* в текстах наших 36 авторов (7054,74) близко к *ipm* «цветных» прилагательных по поэтическому корпусу НКРЯ в целом (7776,48), что является косвенным доказательством репрезентативности нашего корпуса и позволяет говорить о достоверности полученных результатов.

Достоверно выявлено, что в поэтических текстах наблюдается более высокая частотность прилагательных цвета, чем в текстах других типов. Выборочный просмотр результатов поиска в поэтическом корпусе показал вероятность того, что в поэтическом языке преобладает употребление изучаемых прилагательных в метафорическом значении. Изложенная выше проблема разделения значений не позволяет количественно подтвердить это утверждение.

Обращение к базам данных художественных музеев (не только Эрмитажа, но и других), дает некоторые дополнительные возможности, поскольку в описаниях художественных произведений прилагательные цвета употребляются почти исключительно в прямом значении.

Список литературы / References

- Василевич А. П., Кузнецова С. Н., Мищенко С. С.** Цвет и названия цвета в русском языке / Под общ. ред. А. П. Василевича. М.: КомКнига, 2005. 216 с.
Vasilevich, A. P., Kuznecova, S. N., Mishchenko. S. S. Color and color names in the Russian language. Ed. by A. P. Vasilevich. Moscow, KomKniga Publ., 2005, 216 p. (in Russ.)
- Герасимов Г. И.** Прилагательные, обозначающие цвет // Русский язык за рубежом. 1969. № 3. С. 26–35.
Gerasimov, G. I. Adjectives denoting color. *Russkij yazyk za rubezhom* [Russian Language Abroad], 1969, no. 3, p. 26–35. (in Russ.)
- Герасимов Г. И.** Снежно-белый, изумрудно-зеленый: о сложных прилагательных, обозначающих цвет // Русская речь. 1978. № 2. С. 79–84.
Gerasimov, G. I. Snow-white, emerald-green: on complex adjectives denoting color. *Russkaya rech'* [Russian Speech], 1978, no. 2, p. 79–84. (in Russ.)
- Гришина Е. А., Корчагин К. М., Плунгян В. А., Сичинава Д. В.** Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 71–113.
Grishina, E. A., Korchagin, K. M., Plungyan, V. A., Sichinava, D. V. The poetic corpus within the framework of the RNC: General structure and prospects of use. In: *Nacional'ny korpus russkogo yazyka: 2006–2008. Noveye rezul'taty i perspektivy* [National corpus of the Russian language: 2006–2008. New results and prospects]. St. Petersburg, 2009, p. 71–113. (in Russ.)
- Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В.** Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. М., 2005. С. 155–174.
Kustova, G. I., Lyashevskaya, O. N., Paducheva, E. V., Rakhilina, E. V. Semantic markup of vocabulary in the national corpus of the Russian language: principles, problems, prospects. In: *Nacional'ny korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektivy* [The National corpus of the Russian language: 2003–2005. Results and prospects]. Moscow, 2005, p. 155–174. (in Russ.)
- Ляшевская О. Н., Шаров С. А.** Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009. 1087 с.
Lyashevskaya, O. N., Sharov, S. A. The Frequency dictionary of the modern Russian language (the Russian National Corpus). Moscow, Azbukovnik Publ., 2009, 1087 p. (in Russ.)

- Чекменева А. В.** Оригинальные метафорические употребления цветообозначений (корпусный анализ) // Компьютерная лингвистика и вычислительные онтологии (Тр. XXI Международной объединенной конференции «Интернет и современное общество», IMS-2018): Сб. науч. ст. СПб., 2018. Вып. 2. С. 131–147.
- Chekmeneva, A. V.** Original metaphorical use of color terms (corpus analysis). In: *Komp'yuternaya lingvistika i vychislitel'nye ontologii* (Trudy XXI Mezhdunarodnoj ob"edinennoj konferencii «Internet i sovremennoe obshchestvo, IMS-2018) [Computational linguistics and computational ontologies (Proceedings of the XXI International Joint Conference "Internet and modern society", IMS-2018)]. St. Petersburg, University ITMO, 2018, issue 2, p. 131–147. (in Russ.)
- Berlin, B., Kay, P.** Basic Color Terms: Their Universality and Evolution. Berkeley, University of California Press, 1969.
- Kay, P., Maffi, L.** Color Appearance and the Emergence and Evolution of Basic Color Lexicons. *American Anthropologist*, 1999, vol. 101 (4), p. 743–760.
- Kay P., McDaniel Ch. K.** The Linguistic Significance of Meanings of Basic Color Terms. *Language*, 1978, vol. 54 (3), p. 610–646.
- Kay P., Regier T.** Language, thought and color: recent developments. *Trends in Cognitive Sciences*, 2006, vol. 10 (2), p. 51–54.
- Levinson, Stephen C.** Yeli Dnye and the Theory of Basic Color Terms. *Journal of Linguistic Anthropology*, 2000, vol. 10 (1), p. 3–55.

*Материал поступил в редколлегию
Date of submission
20.09.2018*

Сведения об авторах / Information about the Authors

Масевич Андрей Цезаревич, старший преподаватель кафедры информационного менеджмента Санкт-Петербургского государственного института культуры (Дворцовая наб., 2, Санкт-Петербург, 191186, Россия)

Andrey Ts. Masevich, St. Petersburg State Institute of Culture (2 Dvortsovaya Emb., St. Petersburg, 191186, Russian Federation)
andmasev@mail.ru

Захаров Виктор Павлович, кандидат филологических наук, доцент Санкт-Петербургского университета, ведущий научный сотрудник Института лингвистических исследований РАН (Университетская наб., 7-9, Санкт-Петербург, 199034, Россия)

Victor P. Zakharov, St. Petersburg State University (7–9 Universitetskaya Emb., St. Petersburg, 199034, Russian Federation)

v.zakharov@spbu.ru
ORCID 0000-0003-0522-7469