

¹В.П. Захаров, ²А.Ц. Масевич

¹Санкт-Петербургский государственный университет

²Санкт-Петербургский государственный институт культуры

v.zakharov@spbu.ru, andmasev@mail.ru

(Россия, Санкт-Петербург)

КАТЕГОРИЗАЦИЯ ПРИЛАГАТЕЛЬНЫХ ЦВЕТА В РУССКИХ ПОЭТИЧЕСКИХ ТЕКСТАХ (КОРПУСНОЕ ИССЛЕДОВАНИЕ)

Статья описывает некоторые количественные данные использования прилагательных, обозначающих цвета в русских поэтических текстах. Предложена методика выявления прилагательных цвета из текстов поэтического корпуса НКРЯ посредством использования тегов семантической разметки. Произведено сравнение частотности прилагательных цвета в четырех корпусах НКРЯ, при этом выявлено, что частота прилагательных цвета в текстах поэтического корпуса значительно выше, чем в текстах остальных трех корпусов. На основе анализа 170 прилагательных цвета, извлеченных из текстов поэтического корпуса, произведена категоризация полученного списка прилагательных. Приведены количественные данные о представленности (репрезентации) выделенных типов прилагательных в русских поэтических текстах. Выявлена заметная корреляция частоты встречаемости 11 прилагательных, означающих базовые цветовые категории, с эволюционной теорией базовых цветовых терминов Берлина-Кэя.

Ключевые слова: прилагательные цвета, теория базовых цветовых терминов Берлина-Кэя, Национальный корпус русского языка, поэтический корпус, семантическая разметка.

Введение

Обозначение цвета в языке — исключительно популярная тема в лингвистических исследованиях. В данной работе мы пытаемся ввести в рассмотрение несколько принципиально новых аспектов лингвистической статистики с применением лингвистических корпусов.

Цели исследования:

1) выявление возможной корреляции между эволюционной теорией базовых цветовых терминов Брента Берлина и Пола Кэя (Brent Berlin & Paul Kay) [Berlin, Kay 1969] и статистикой частоты встречаемости прилагательных цвета;

2) изучение частотного поведения единиц данной лексической группы на материале поэтических текстов в сравнении с другими типами текстов;

3) систематизация прилагательных цвета.

Исследование выполнялось по следующему плану:

1. Выявление и анализ частотности прилагательных цвета в четырех корпусах Национального корпуса русского языка (НКРЯ).

2. Создание авторских подкорпусов для отдельных авторов.

3. Выявление количества прилагательных цвета, встречающихся в текстах отдельных авторов.

4. Систематизация прилагательных цвета на основе полученных данных.

В качестве материала исследования использовались:

1) поэтический корпус НКРЯ и авторские корпуса, созданные на его основе;

2) корпуса поэтических текстов, созданные специально для целей исследования в системе Sketch Engine на базе различных источников;

3) основной, газетный и устный корпуса НКРЯ.

В качестве инструментов исследования использовались:

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

2) корпусный менеджер Sketch Engine;

3) электронная таблица MS Excel 2010.

Теория базовых цветовых терминов Берлина и Кэя

Теория базовых цветовых терминов (категорий) Б. Берлина и П. Кэя состоит в том, что языки мира содержат полностью или частично общий набор слов цветообозначений, и эти слова в процессе исторического развития появляются в языках в определенном порядке. В английском языке авторы выделяют 11 базовых прилагательных цвета, в славянских, в частности в русском, таких прилагательных 12.

Исследовав десятки европейских языков, с одной стороны, и десятки языков целого ряда примитивных культур — с другой, Берлин и Кэй открыли «всеобщий» эволюционный закон. По их мнению, существует семь ступеней (стадий) развития цветовой терминологии, отражающих последовательность появления в лексике языка каждого такого слова [Berlin, Kay 1969; Kay, McDaniel 1978; Kay, Regier 2006].

Критерии принадлежности к базовым категориям следующие:

1) слово должно быть непроемным 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).

На «самой низшей» стадии, стадии I, в языке есть только два основных цветообозначения, и это во всех случаях оказываются слова, указывающие на белый и черный цвета. На стадии II к двум цветам добавляется третий, и это всегда слово,

обозначающее красный цвет. На стадиях III—V добавляется каждый раз по одному слову из трех — «синий», «зеленый», «желтый». Седьмым словом (стадия VI) всегда бывает «коричневый», а высшая стадия VII характеризуется появлением четырех цветов — «розовый», «оранжевый», «фиолетовый» и «серый» [Василевич, Кузнецова, Мищенко 2005].

Анализ прилагательных цвета по четырем корпусам НКРЯ

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. При поиске семантические признаки задаются с помощью вкладок — меню, которые раскрываются по ссылке на входной странице поиска «Семантические признаки: выбрать». Существует восемь меню с перечнем семантических признаков для пяти частей речи. Существительным соответствуют три меню — для имен существительных предметных, непредметных и собственных. Семантический признак «цвет» может быть задан для трех частей речи — прилагательных, непредметных существительных и наречий (рис. 1).

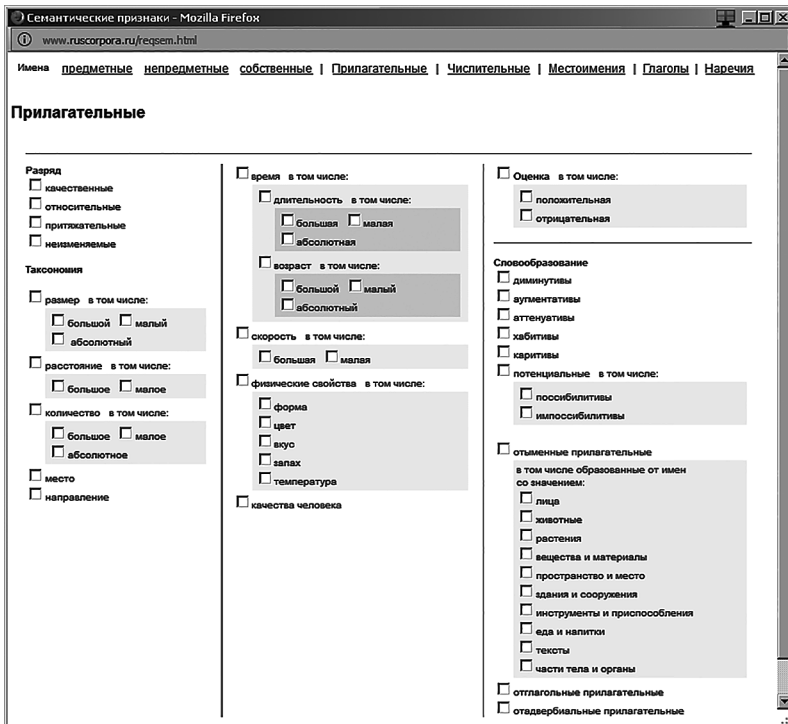


Рис. 1. Меню выбора семантических признаков для прилагательных

Таблица 1

Результат поиска по семантическому признаку «цвет» в четырех корпусах

Корпус	Число словоупотреблений (вхождений)	ipm
Поэтический корпус	85286	7776,48
Основной корпус	771638	2722,48
Газетный корпус	280630	1228,02
Устный корпус	12682	1046,93

На первом этапе мы отбирали все контексты с прилагательными цвета в четырех корпусах: поэтическом, основном, газетном и устном, — опираясь именно на семантический тег «Физические свойства: цвет». Результаты представлены в таблице 1.

Анализ данных таблицы показывает, что значение ipm прилагательных, означающих цвет, в поэтическом корпусе в 2,9 раза выше, чем в основном корпусе, в 6,3 раза выше, чем в газетном корпусе, и в 7,4 раза выше, чем в устном корпусе.

Для сопоставления частотности цветообозначающих прилагательных отобраны 11 прилагательных, которые в известных экспериментах Берлина и Кэя назывались базовыми цветовыми категориями.

Результаты поиска по этим прилагательным по четырем корпусам приведены в таблице 2. Кроме того, в таблице приводится частотность отобранных прилагательных по Частотному словарю современного русского языка (ЧССРЯ) [Ляшевская, Шаров 2009].

Таблица 2

Частотность 11 базовых цветовых категорий в четырех корпусах и в Частотном словаре

Прилагательное	Поэтический корпус		Основной корпус		Газетный корпус		Устный корпус		ЧССРЯ
	словоупотр.	ipm	словоупотр.	ipm	словоупотр.	ipm	словоупотр.	ipm	
белый	10730	978,37	111109	392,01	54079	236,65	2115	174,60	339,6
черный	9556	871,33	106635	376,23	42210	184,71	2286	188,72	337,5
красный	5582	508,97	86762	306,11	47301	206,99	1958	161,64	240,5
синий	4951	451,44	29371	103,63	8722	38,17	450	37,15	82,8
зеленый	4871	444,14	37891	133,69	17006	74,42	854	70,50	124,4
серый	2692	245,46	38297	135,12	7753	33,93	453	37,40	96,1
желтый	2423	220,93	26833	94,67	8445	36,95	388	32,03	78,1
розовый	1883	171,69	16436	57,99	3938	17,23	216	17,83	49,3
оранжевый	260	23,71	3944	13,92	6825	29,87	67	5,53	15,60
коричневый	181	16,50	6960	24,56	1491	6,52	110	9,08	25,5
фиолетовый	125	11,40	3325	11,73	727	3,18	57	4,71	12,4
среднее		358,54		149,97		78,96		67,20	127,44

Мы видим, что частота встречаемости прилагательных цвета в поэтическом корпусе заметно выше, чем их частота в других корпусах. При этом разница в ipm прилагательных цвета между поэтическим и основным корпусом меньше, чем между поэтическим и газетным и устным корпусами. Это легко объясняется тем, что

основной корпус сбалансирован по жанровому составу. Из таблицы 2 видно, что значение *irm* базовых категорий в основном корпусе близко к значениям ЧССРЯ.

Для 11 прилагательных цвета статистическая картина встречаемости их в четырех корпусах НКРЯ в целом повторяет данные, полученные при поиске по семантическому тегу «цвет» (см. табл. 1), что хорошо видно как по отдельным цветам, так по средним значениям *irm*. Лишь в двух случаях (прилагательные «коричневый» и «фиолетовый») *irm* в основном корпусе незначительно превышают *irm* в поэтическом корпусе, и в одном случае в газетном корпусе *irm* прилагательного «оранжевый» превышает частотность этого же прилагательного в поэтическом корпусе.

Обратим внимание на некоторые различия в порядке снижения частотности прилагательных. Для этого расположим слова в порядке убывания частоты для каждого корпуса (табл. 3). Под таблицей расположим эволюционную схему Берлина-Кэя.

Таблица 3

Порядок убывания частоты прилагательных цвета в четырех корпусах НКРЯ в сопоставлении с эволюционной схемой Берлина-Кэя

Корпус Номер	1	2	3	4	5	6	7	8	9	10	11
Поэтический	белый	черный	красный	синий	зеленый	серый	желтый	розовый	оранжевый	коричневый	фиолетовый
Основной	белый	черный	красный	серый	зеленый	синий	желтый	розовый	коричневый	оранжевый	фиолетовый
Газетный	белый	красный	черный	зеленый	синий	желтый	серый	оранжевый	розовый	коричневый	фиолетовый
Устный	черный	белый	красный	зеленый	серый	синий	желтый	розовый	коричневый	оранжевый	фиолетовый

Берлин-Кэй	Стадия I		Стадия II	Стадия III—IV		Стадия V	Стадия VI	Стадия VII			
Берлин-Кэй	белый	черный	красный	зеленый	желтый	синий	коричневый	серый	розовый	оранжевый	фиолетовый

Как видно из таблицы 3, из 11 отобранных прилагательных в трех корпусах самым частотным оказывается «белый», в одном случае (устный корпус) — «черный». Во всех четырех корпусах три прилагательных «белый», «черный» и «красный» составляют тройку самых частотных. Следующие по частоте — четыре прилагательных «синий», «серый», «зеленый» и «желтый». Наименее частотными оказываются прилагательные «розовый», «оранжевый», «коричневый» и «фиолетовый». Порядок следования этих прилагательных по убыванию частоты в целом коррелирует со стадиями появления слов-цветообозначений в различных языках

по Берлину-Кэю, хотя прямо не совпадает с ними. Например, прилагательное «желтый» при полной корреляции должно было бы быть более частотным, чем прилагательное «синий», а по данным НКРЯ его частотность ниже «синего». Также прилагательное «серый», относящееся к седьмой, последней, исторической стадии, по нашим данным имеет сравнительно высокую частотность и соответствует стадиям IV и V.

Еще два наблюдения. Во-первых, порядок убывания частотности во всех корпусах различается не очень существенно. Можно, как описано выше, разбить массив из 11 прилагательных на три группы — в первой три прилагательных, в двух других по четыре. Обращает на себя внимание, что в пределах одной группы порядок убывания частотности у прилагательных может различаться, но выявлен всего один случай «выхода» за пределы группы. Во-вторых, частотность прилагательных «белый», «черный» и «красный» почти полностью коррелирует с первыми двумя стадиями эволюционной схемы.

Создание авторских корпусов

Корпус поэтических текстов включает стихотворные произведения от XVIII в. до современности. Помимо обычной семантической и морфологической разметки, имеется специальная стиховедческая разметка [Гришина и др. 2009]. Для создания авторских корпусов были отобраны наиболее известные поэты, представляющие различные хронологические периоды. При этом выяснилось, что тексты некоторых показательных авторов, в основном второй половины XX в., в НКРЯ отсутствуют. В связи с этим авторские корпуса этих поэтов были созданы с помощью системы Sketch Engine на основе текстов, представленных на авторитетных веб-сайтах.

Таблица 4

Список авторов с указанием источников текстов для создания авторских корпусов

Автор	Источник
1. Аронзон	НКРЯ
2. Асадов	Интернет
3. Ахмадулина	Интернет
4. Ахматова	НКРЯ
5. Бальмонт	НКРЯ
6. Баратынский	НКРЯ
7. Блок	НКРЯ
8. Бродский	Интернет
9. Вертинский	НКРЯ
10. Вознесенский Андрей	Интернет

Автор	Источник
11. Высоцкий	Интернет
12. Державин	НКРЯ
13. Евтушенко	Интернет
14. Есенин	НКРЯ
15. Жданов	НКРЯ
16. Заболоцкий	НКРЯ
17. Кирилов	Интернет
18. Кривулин	НКРЯ
19. Лермонтов	НКРЯ
20. Мандельштам	НКРЯ

Автор	Источник
21. Матвеева Новелла	Интернет
22. Маяковский	НКРЯ
23. Набоков	НКРЯ
24. Некрасов	НКРЯ
25. Пастернак	НКРЯ
26. Пушкин	НКРЯ
27. Рубцов	НКРЯ
28. Светлов	НКРЯ

Автор	Источник
29. Симонов	НКРЯ
30. Слуцкий	НКРЯ
31. Тарковский	НКРЯ
32. Твардовский	НКРЯ
33. Толстой А. К.	НКРЯ
34. Тютчев	НКРЯ
35. Хлебников	НКРЯ
36. Цветаева	НКРЯ

Методика исследования

Создание таблиц частотности прилагательных цвета для отдельных авторов средствами НКРЯ

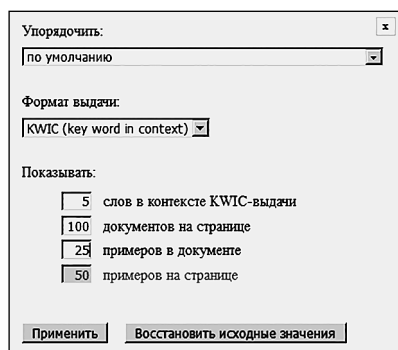


Рис. 2. Параметры представления результатов при поиске в авторском корпусе НКРЯ

По созданному авторскому корпусу выполняется запрос на прилагательные с семантическим признаком «цвет». В НКРЯ это осуществляется посредством вкладки, описанной выше (см. рис.1). Следующий шаг — статистическая обработка полученных результатов. Идеальный вариант — получить сразу всю статистику по данному запросу — как правило, не проходит из-за ограничений на объем выдачи (рис. 2). Опытным путем мы установили, что максимальное число документов, отражаемых на одной странице, не должно превышать 100. Число примеров в одном документе мы ограничивали числом 25 (см. рис. 2).

В нижней части каждой страницы с результатами поиска находится раздел, который называется «Частоты найденного для этой страницы» (рис. 3).

С каждой страницы поиска производится перенос всех таблиц «Леммы» в MS Excel, где формируется сводная таблица на весь результат поиска по данному авторскому подкорпусу. Средствами MS Excel осуществляется подсчет числа употреблений каждого прилагательного, его доли в процентах среди других прилагательных цветообозначения, а также относительной частоты (ipm) по отношению к числу словоупотреблений во всем авторском подкорпусе.

Формула подсчета доли данного прилагательного среди всех прилагательных цвета в данном подкорпусе следующая:

Частоты найденного для этой страницы

Словоформы			Леммы		
1	яркие	3	1	яркий	4
2	яркими	1	2	красный	4
3	холодном	1	3	синий	2
4	теплым	1	4	светлым	2
5	темным	1	5	лимонный	2
6	синим	1	6	золотой	2
7	синее	1	7	холодный	1
8	серые	1	8	теплым	1
9	светлому	1	9	темным	1
10	светлое	1	10	серый	1
11	розовыми	1	11	розовый	1
12	лимонный	1	12	зеленый	1
13	лимонного	1			
14	красным	1			
15	красный	1			
16	красное	1			
17	класное	1			
18	золотым	1			
19	золотое	1			
20	зеленый	1			

Рис. 3. Статистика запроса — частоты лемм и словоформ

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
19	мутный	Блок	11	96,92	0,67
20	златой	Блок	8	70,49	0,49
21	рыжий	Блок	8	70,49	0,49
22	серебристый	Блок	8	70,49	0,49

Рис. 4. Фрагмент таблицы MS Excel частотности прилагательных, обозначающих цвет в авторском подкорпусе текстов А. А. Блока

$$N/B * 100,$$

где N — число словоупотреблений данной леммы в исследуемом авторском корпусе, B — число словоупотреблений прилагательных, обозначающих цвета в результатах поиска по авторскому корпусу.

Формула подсчета ipm следующая:

$$N/A * 1000000,$$

где N — число словоупотреблений данной леммы в исследуемом авторском корпусе,

A — общее число словоупотреблений в исследуемом авторском корпусе.

В результате описанных операций создается таблица частотности прилагательных цвета в корпусе данного автора (рис. 4).

Создание таблиц частотности прилагательных цвета для отдельных авторов средствами Sketch Engine

В системе Sketch Engine нет семантической разметки, поэтому методика обработки данных отличалась от описанной выше для корпусов НКРЯ. Запрос задавался в явном виде, а именно перечислением всех прилагательных цвета, полученных на основе данных НКРЯ. В связи с ограничением на длину запроса проводилось два сеанса поиска: по простым (однокоренным) (рис. 5) и сложным (двусоставным) (рис. 6) прилагательным.

агатовый | аленький | алый | аметистовый | багровый | багряный | бежевый | белейший | беленький | белесоватый | белесый | беловатый | белоснежный | белый | бесцветный | бирюзовый | бледный | бордо | бордовый | буланный | бурый | вороной | гнедой | голубенький | голубоватый | голубой | двухцветный | дымчатый | желтенький | желтый | зелененький | зеленоватый | зеленый | золотой | золотистый | изумрудный | караковый | карий | кауры | клетчатый | коричневый | красненький | красноватый | красный | лазоревый | лазурный | лиловатый | лиловый | линиялый | мутный | одноцветный | опаловый | оранжевый | палевый | пегий | пестрый | поблеклый | пожелтелый | поседель | посинелый | потускнелый | почернелый | пунцовый | пурпурный | пурпуровый | пятнистый | пятнистый | разноцветный | розоватый | розовый | румяный | русый | рыжебородый | рыжеватый | рыжеволосый | рыженький | рыжий | рябенький | рябоватый | саврасый | самоцветный | светлый | седенький | серебристый | серенький | сероватый | серый | сивый | сизокрылый | сизый | синеватый | синенький | синий | соловый | сумрачный | сумрачный | темнехонький | темный | трехцветный | фиолетовый | цветастый | цветной | чалый | червленький | черненький | чернехонький | чернешенький | черноватый | черноволосый | черномазый | черный

Рис. 5. Простые (однокоренные) прилагательные цвета

.*-красный|.*-оранжевый|.*-желтый|.*-зеленый|.*-голубой|.*-синий|.*-фиолетовый|.*-коричневый|.*-белый|.*-черный|.*-розовый|.*-лиловый|светло-.*|темно-.*|ярко-.*|черно-.*|сине-.*|нежно-.*|серо-.*

Рис. 6. Двусоставные прилагательные цвета (с переменной первой или второй частью)

По результатам автоматически, с использованием функции Frequency, формировался частотный список лемм, затем сводный список по двум сеансам поиска. Из сводного списка удалялись все слова, не имеющие значения «цвет» (например, «ярко-тугой», «ярко-положительный», «нежно-хмельной»). Все нелемматизированные формы (например, «ярко-золотистым», «ярко-желтым», «ярко-розовою», «черно-золотая») приводились к лемме. Обработанный таким образом список переносился в таблицу Excel в соответствующие столбцы («Лемма» и «Число с/у»), где далее формировалась сводная таблица на весь результат поиска по данному авторскому подкорпусу, как это описано выше.

Исследование соотношений частоты употреблений прилагательных цвета по текстам 36 авторов

Полученные таблицы по каждому из 36 авторов были слиты в единую таблицу Excel, составившую 1571 строку. Материал в таблице отсортирован по столбцу «Лемма». Всего в таблице оказалось 170 лемм, и средствами Excel было подсчитано количество авторов, у которых каждая из них встретилась. В таблице 5 показаны первые (наиболее частотные) 25 прилагательных из 170.

Таблица 5

Наиболее часто встречающиеся прилагательные, обозначающие цвета

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. белый	36	10. бледный	35	19. белоснежный	28
2. голубой	36	11. серый	34	20. золотистый	26
3. желтый	36	12. розовый	33	21. лиловый	24
4. зеленый	36	13. рыжий	31	22. сизый	24
5. красный	36	14. цветной	31	23. сумрачный	24
6. светлый	36	15. багровый	30	24. серебристый	23
7. синий	36	16. пестрый	30	25. багряный	21
8. темный	36	17. алый	29		
9. черный	36	18. мутный	29		

Категоризация прилагательных цвета

Сделана попытка разбить 170 прилагательных на классы, иначе говоря, разработать основы их категоризации. Естественно, состав этих классов открыт, а возможно, появятся (следует ввести) и какие-то другие.

Первая группа — собственно цвета, таких слов мы обнаружили 30 из 170 (17,6%) (табл. 6).

Таблица 6

Прилагательные, означающие собственно цвета

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. белый	36	12. лиловый	24	23. пурпуровый	10
2. голубой	36	13. сизый	24	24. сивый	10
3. желтый	36	14. багряный	21	25. палевый	9
4. зеленый	36	15. лазурный	20	26. бордовый	5
5. красный	36	16. бурый	18	27. червлённый	5
6. синий	36	17. оранжевый	18	28. бежевый	3
7. черный	36	18. пурпурный	16	29. бордо	3
8. серый	34	19. коричневый	13	30. опаловый	3
9. розовый	33	20. лазоревый	13	Среднее	20,5
10. багровый	30	21. фиолетовый	12		
11. алый	29	22. пунцовый	10		

В эту группу вошли все 11 основных цветов (12 для русского языка). В пределах одного частотного ранга слова отсортированы по алфавиту. Отметим, что и здесь частотность (число авторов) у прилагательных «серый», «розовый», «оранжевый»,

«коричневый» и «фиолетовый», соответствующих VI и VII эволюционным ступеням, ниже частотности прилагательных, относящихся к более ранним ступеням, что еще раз говорит о корреляции между частотностью слов в поэтических текстах и стадиями эволюции по Берлину и Кэю.

Среди этих прилагательных встречаются синонимы. Вряд ли можно сделать уверенное различие между значениями прилагательных «красный», «алый», «пунцовый», «червлёный» или «фиолетовый», «лиловый», «пурпуровый», а также, например, «бурый» и «коричневый».

Кроме того, пожалуй, нет ни одного слова из тех, что относятся к базовым категориям цветообозначений, которое бы не употреблялось в переносном смысле — метафорически либо символически — в составе устойчивых словосочетаний. Достаточно вспомнить выражения «белый свет», «черный список», «красная нить», «оранжевая революция», «желтая пресса», «тоска зеленая», «голубой» (о мужчине нетрадиционной сексуальной ориентации), «синий чулок», «коричневая чума».

Вторую группу слов мы условно назвали «квазицвета» (19 лемм, 11,2%) (табл. 7). В число прилагательных, обозначающих цвета (то есть имеющих тег «цвет» в НКРЯ), попадают слова «темный», «светлый», «бледный», «пестрый», «сумрачный», «цветной», которые, собственно, означают не цвет, а его интенсивность, оттенок (темный, светлый, бледный), просто наличие цвета (цветной) — как оппозиция черному и белому, множество цветов (пестрый) или даже настроение (сумрачный), которое может вызвать цвет или комбинация цветов.

Таблица 7

Прилагательные «квазицвета»

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. светлый	36	8. разноцветный	21	15. трехцветный	3
2. темный	36	9. бесцветный	20	16. двухцветный	1
3. бледный	35	10. клетчатый	10	17. крапчатый	1
4. цветной	31	11. самоцветный	10	18. одноцветный	1
5. пестрый	30	12. пятнистый	9	19. потускнелый	1
6. мутный	29	13. цветастый	8	среднее	16,42
7. сумрачный	24	14. линиялый	6		

С одной стороны, вызывает сомнение приписывание этим прилагательным тега «цвет». Они описывают визуальные характеристики объекта, но не цвет. С другой стороны, многие из этих слов хотя и не обозначают, но характеризуют именно цвет. Некоторые из них образуют сложные «цветные» прилагательные типа «бледно-зеленый», «светло-серый», «темно-красный» и т. д. (см. ниже табл. 11).

Как видно из таблицы 7, употребляемость некоторых из этих прилагательных достаточно высока. Особенно это касается прилагательных «светлый» и «темный».

Отдельные авторы утверждают, что именно смена светлого и темного времени суток обусловила появление цветообозначений «белый» и «черный» на самой ранней стадии эволюции цветообозначений в языках.

Третью группу составляют прилагательные, которые «привязываются» к определенному, чаще одушевленному, предмету. Это может быть цвет волос или глаз, масть или окрас животных (28 лемм, 16,5%) (табл. 8).

Таблица 8

Прилагательные, привязанные к определенному объекту

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. рыжий	31	11. сизокрылый	7	21. седоватый	2
2. румяный	21	12. буланный	6	22. соловый	2
3. русый	21	13. порыжелый	4	23. каурый	1
4. гнедой	15	14. рыжеволосый	4	24. муругий	1
5. карий	14	15. рыженький	4	25. мухортый	1
6. пегий	10	16. рябоватый	3	26. рыжебородый	1
7. вороной	9	17. саврасый	3	27. рябенький	1
8. поседелый	9	18. чалый	3	28. чубарый	1
9. рыжеватый	7	19. черноволосый	3	среднее	6,93
10. седенький	7	20. черномазый	3		

Некоторая часть этих слов также не означает собственно цвета. В некоторых случаях они означают цветовую неоднородность, наличие пятен («пегий», «рябенький», «чубарый», «чалый»), в других — указывается на результат некоторого процесса («поседелый», «порыжелый»). В некоторых случаях образуется прилагательное, содержащее одновременно объект и цветовую характеристику («рыжебородый», «черноволосый»). Значительное место в этой группе занимают масти лошадей и окрасы собак. Известно, что масть или окрас это не столько цвет, сколько сочетание цветов, генетически детерминированный тип распределения пигментов в шерсти животного.

Четвертая группа (85 лемм, 50%) объединяет прилагательные-дериваты. Она может быть разделена на три подгруппы, отражающие словообразовательные модели образования прилагательных.

В первую подгруппу четвертой группы вошли прилагательные-диминывы (13 лемм, 7,6%) (табл. 9).

В их число вошли не только собственно прилагательные цвета — мы видим среди них и прилагательные «темнехонький» и «бледнецкий», исходные формы которых «темный» и «бледный» мы отнесли к «квазичветам». Некоторые из них мы включили в таблицу 7: «рыженький», «рябенький» и «седенький». Здесь мы сталкиваемся с типичной проблемой любой классификации — два значимых классификационных признака у одного объекта.

Таблица 9

Прилагательные-димиутивы

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. беленький	13	6. черненький	5	11. темнехонький	1
2. серенький	13	7. зелененький	3	12. чернехонький	1
3. синенький	9	8. желтенький	2	13. чернешенький	1
4. аленький	6	9. красненький	2	среднее	4,77
5. голубенький	5	10. бледненький	1		

Вторая подгруппа — это прилагательные, которые мы условно назвали прилагательными подобия (10 лемм, 5,9%) (табл. 10).

Таблица 10

Прилагательные подобия

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. розоватый	13	5. голубоватый	7	9. белесоватый	2
2. синеватый	11	6. лиловатый	6	10. беловатый	1
3. зеленоватый	9	7. желтоватый	4	среднее	6,5
4. красноватый	8	8. сероватый	4		

В этой форме мы видим некоторые прилагательные из таблицы 8: «рябоватый», «рыжеватый» и «седоватый». Обращает внимание, что их исходные слова те же, от которых были образованы димиутивы.

Третья подгруппа — это сложные прилагательные, имеющие обычно два (редко три) корня (62 леммы, 36,4%) (табл. 11). Они разными способами описывают интенсивность или оттенок цвета.

Таблица 11

Сложные прилагательные

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. темно-синий	14	8. ярко-желтый	4	15. ярко-красный	3
2. темно-голубой	9	9. молочно-белый	3	16. бледно-зеленый	2
3. черно-белый	9	10. нежно-белый	3	17. бледно-лиловый	2
4. темно-зеленый	8	11. светло-серый	3	18. медно-красный	2
5. бледно-голубой	7	12. светло-синий	3	19. нежно-голубой	2
6. черно-синий	6	13. серо-зеленый	3	20. светло-голубой	2
7. светло-зеленый	5	14. серо-синий	3	21. сине-зеленый	2

Продолжение табл. 11

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
22. синий-синий	2	36. нежно-зеленый	1	50. черно-бархатный	1
23. темно-желтый	2	37. нежно-золотой	1	51. черно-бурый	1
24. темно-красный	2	38. нежно-розовый	1	52. черно-золотой	1
25. темно-серый	2	39. потерто-белый	1	53. черно-красной	1
26. бледно-розовый	1	40. пречерный-черный	1	54. черно-серебряный	1
27. буйно-красный	1	41. прозрачно-синий	1	55. электродово-синий	1
28. грязно-белый	1	42. розовато-белый	1	56. ярко-белые	1
29. грязно-розовый	1	43. светло-пегий	1	57. ярко-голубой	1
30. желто-зеленый	1	44. светло-сиреневый	1	58. ярко-голубой	1
31. зеленовато-голубой	1	45. серо-голубой	1	59. ярко-зеленый	1
32. красно-синий	1	46. сине-красный	1	60. ярко-золотистый	1
33. кроваво-красный	1	47. темно-бирюзовый	1	61. ярко-оранжевый	1
34. медно-зеленый	1	48. темно-вишневый	1	62. ярко-розовый	1
35. небесно-голубой	1	49. темно-лиловый	1	среднее	2,24

Число лемм сложных прилагательных значительно превышает число членов других групп и подгрупп. В том числе элемент «бледно» образует 4 слова, «нежно» — 5 слов, «светло» — 6 слов, «темно» — 9 слов, «ярко» — 9 слов.

Наконец, **пятая группа**, которую мы назвали «аналоговые прилагательные» (табл. 12). Они передают цвета и оттенки через цвета и оттенки хорошо известных предметов. Г. И. Герасимов называет их относительными прилагательными [Герасимов 1969]. Семантическая разметка НКРЯ плохо учитывает слова такого типа, причем многие из них не имеют тега «цвет», поэтому при поиске по семантическому признаку в корпусе найдено всего 8 (4,7%) таких прилагательных. На самом деле их гораздо больше.

Таблица 12

Аналоговые прилагательные

Лемма	Число авторов	Лемма	Число авторов	Лемма	Число авторов
1. золотистый	26	4. изумрудный	18	7. агатовый	3
2. серебристый	23	5. бирюзовый	10	8. аметистовый	2
3. золотой	21	6. дымчатый	8		

Проблема, видимо, в полисемии этих прилагательных. Рассмотрим одно из наиболее характерных прилагательных этого типа — «малиновый». Его описание в корпусе содержит следующие теги семантической разметки:

Семантика основная der:s, dt:food, dt:fruit, r:rel

Семантика дополнительная; ev:posit, r:rel, t:sound.

Эти теги указывают на происхождение слова от существительного «малина», то, что оно относится к еде (тег «food»), тег «fruit» подразумевает понятие «ягода», его оценка позитивная «ev:posit», тег «t:sound» указывает на отношение к звукам (по-видимому, «малиновый звон»), но ни в основной, ни в дополнительной семантике цвета нет.

Ниже приведены 10 самых частотных биграмм с прилагательным «малиновый», полученных с помощью системы Google Books Ngram Viewer.

1996

малиновый цвет_NOUN

малиновый перезвон_NOUN

малиновый звон_NOUN

малиновый кафтан_NOUN

малиновый бархат_NOUN

малиновый закат_NOUN

малиновый оттенок_NOUN

малиновый Евсюков_NOUN

малиновый свет_NOUN

малиновый чай_NOUN

В этом списке видно, что преобладающим значением прилагательного является именно цвет. Даже комиссар Евсюков из рассказа Б. Лавренева «Сорок первый» назван малиновым потому, что носил куртку малинового цвета.

Вывод, казалось бы, ясен. При ближайшем редактировании семантической разметки НКРЯ аналоговым прилагательным следует приписать тег «цвет». Однако все не так просто. Не у каждого прилагательного этого ряда значение цвета является доминирующим. Например, у таких прилагательных, как «кровавый», «стальной», цвет будет, скорее, периферийным значением, в результате чего применение этого тега могло бы вызвать значительный информационный шум.

Заключение

Прилагательные, которые рассмотрены нами в данной статье, имеют один общий признак — наличие тега «цвет» в семантической разметке НКРЯ. По приведенным данным частота встречаемости прилагательных цвета в поэтическом корпусе выше, чем их частота в основном корпусе, и заметно выше, чем в устных речевых произведениях или газетных статьях.

Прилагательные цвета могут быть разделены на пять групп. Причем четвертая группа (формы прилагательного) может быть, в свою очередь, разбита на три подгруппы. В ходе исследования выявлены некоторые недостатки семантической разметки, которые не позволили в нашем случае учесть истинное число лемм прилагательных, которые мы назвали аналоговыми.

Что касается частотного поведения прилагательных цвета, то здесь можно обозначить два аспекта. С одной стороны, имеется корреляция частоты прилагательных с эволюционной теорией Берлина и Кэя, то есть наиболее частотны базовые категории прилагательных, а среди них наиболее частотны прилагательные, относящиеся к более ранней стадии эволюции. С другой стороны, эта корреляция не является абсолютной, но зависит от контекста. При этом отметим, что схема

Берлина и Кэя, как, впрочем, и любая логическая схема, в применении к языку не может быть соблюдена полностью, тем более, что она представляет собой некоторый усредненный результат анализа обширного лингвистического материала.

Употребляемость авторами поэтического корпуса прилагательных собственно цвета выше частотности прилагательных других обозначенных нами групп.

Обращает на себя внимание, что наряду с небольшим числом наиболее частотных прилагательных существует очень большое число сложных цветовых терминов единичного употребления. Подобная структура лексического массива напоминает частотное распределение по закону Ципфа.

Литература

Василевич А. П., Кузнецова С. Н., Мищенко С. С. Цвет и названия цвета в русском языке / под общ. ред. А. П. Василевича. М.: Ком Книга, 2005. 216 с.

Герасимов Г. И. Прилагательные, обозначающие цвет // Русский язык за рубежом. 1969. №3. С. 26–35.

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.

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.

Berlin B., Kay P. Basic Color Terms: Their Universality and Evolution. Berkeley: University of California Press, 1969.

Kay P., Regier T. Language, thought and color: recent developments // Trends in cognitive sciences. 2006. No.10 (2). P. 51–54.

Kay P., McDaniel Ch.K. The Linguistic Significance of Meanings of Basic Color Terms // Language. 1978. No. 54 (3). P. 610–646.

¹*V.P. Zakharov, ²A.Ts. Masevitch*

¹ *St. Petersburg State St. University*
v.zakharov@spbu.ru

² *St. Petersburg State institute of culture*
andmasev@mail.ru

^{1, 2} *(Russia, St. Petersburg)*

CATEGORIZATION OF ADJECTIVES OF COLOR IN THE RUSSIAN POETICAL TEXTS (A CORPUS-BASED STUDY)

The paper describes certain quantitative data of usage of color adjectives in Russian poetic texts. We suggest methodology of revealing adjectives of color in the poetic

corpus of the Russian National Corpus with the help of semantic tags. We have also compared frequency of adjectives of color in four corpora of the Russian National Corpus and found that frequency of adjectives of color in the poetic corpus is significantly higher than in other three corpora. Having analyzed 170 lemmas of adjectives of color retrieved from the texts of the poetic corpus, we have systematized these adjectives and suggested criteria of their typification. The paper presents quantitative data on the representation of suggested types of adjectives in Russian poetic texts. We have revealed a noticeable correlation between the frequency of eleven adjectives of color and evolutionary theory of basic color terms by Berlin and Kay.

Keywords: adjectives of color, theory of basic color terms by Berlin & Kay, Russian National Corpus, poetic corpus, semantic tagging.

References

Berlin B., Kay P. *Basic Color Terms: Their Universality and Evolution*. Berkeley, University of California Press, 1969.

Gerasimov G. I. [Adjectives denoting colour]. *Russkii yazyk za rubezhom* [Russian abroad]. 1969, no. 3, pp. 26–35. (In Russ.)

Grishina E. A., Korchagin K. M., Plungyan V. A., Sichinava D. V. [Poetic corpus within the framework of the RNC: general structure and prospects of use]. *Natsional'nyi korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy* [The Russian National Corpus: 2006–2008. New results and prospects]. St. Petersburg, Nestor-Istoriya Publ., 2009, pp. 71–113. (In Russ.)

Kay P., McDaniel Ch.K. The Linguistic Significance of Meanings of Basic Color Terms. *Language*, 1978, no. 54 (3), pp. 610–646.

Kay P., Regier T. Language, thought and color: recent developments. *Trends in cognitive sciences*, 2006, no. 10(2), pp. 51–54.

Kustova G. I., Lyashevskaya O. N., Paducheva E. V., Rahilina E. V. [Semantic annotation of the lexicon in the Russian National Corpus: principles, problems, prospects]. *Natsional'nyi korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektivy* [The Russian National Corpus: 2003–2005. Results and Prospects]. Moscow, 2005, pp. 155–174. (In Russ.)

Vasilevich A. P., Kuznecova S. N., Mishchenko S. S. *Tsvet i nazvaniya tsveta v russkom yazyke* [Color and colour names in Russian]. Ed. A. P. Vasilevich. Moscow, Kom Kniga Publ., 2005. 216 p. (In Russ.)