

Научная статья

УДК 81-13

DOI 10.25205/1818-7935-2021-19-4-5-35

Диахронические изменения частотности цветообозначений в русских поэтических текстах

Андрей Цезаревич Масевич¹

Виктор Павлович Захаров²

¹ Санкт-Петербургский государственный институт культуры
Санкт-Петербург, Россия

² Санкт-Петербургский государственный университет
Санкт-Петербург, Россия

¹ andmasev@mail.ru, <https://orcid.org/0000-0002-1752-8915>

² v.zakharov@spbu.ru, <https://orcid.org/0000-0003-0522-7469>

Аннотация

С помощью методов корпусной лингвистики и с использованием семантической разметки Национального корпуса русского языка (НКРЯ) выявлены некоторые закономерности изменения частотности прилагательных-цветообозначений в поэтических текстах с 1750 по 1999 г. Попутно приводятся описание и характеристики поэтического корпуса НКРЯ (временной охват, число авторов, текстов, число словоупотреблений с разбивкой по десятилетиям), существенные для оценки его репрезентативности. В ходе исследования также выявлен существенный дефект семантической разметки НКРЯ – непоследовательность присвоения тега *цвет*. На взгляд авторов, этот дефект мог повлиять на частные результаты исследования, однако выявленные тенденции частотного поведения лексических единиц представляются достоверными. Методика исследования основывается на сравнении статистических показателей для каждого десятилетия в указанном диапазоне. Соответственно, графики строятся по средним значениям для диапазона, а не по точкам для каждого года с коэффициентом сглаживания, как в работе с инструментом «Графики» НКРЯ. Также показатели в поэтическом корпусе сравниваются с соответствующими данными в корпусе прозы. Кратко представлена обновленная по сравнению с предложенной в работе А. Ц. Масевича и В. П. Захарова «Частотное поведение прилагательных цвета в русских поэтических текстах» (Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 21–48) классификация прилагательных, имеющих тег *цвет* семантической разметки в НКРЯ. Результаты исследования представлены в виде диаграмм изменения частотности единиц на протяжении XVIII, XIX и XX вв., построенных для повышения их надежности двумя методами. Достоверно установлено, что в поэтических текстах частотность цветообозначений выше, чем в прозе любого жанра независимо от того, к какой части речи относятся слова-цветообозначения. Самые высокие значения частоты (IPM) всех прилагательных, имеющих тег *цвет* семантической разметки НКРЯ, в поэтическом корпусе НКРЯ отмечаются в последние десятилетия XIX – первые десятилетия XX в. Частотное поведение прилагательных, относящихся к разным классам, существенно отличается в разные временные периоды. В поэзии XVIII в. преобладают аналоговые прилагательные (т. е. слова вторичной номинации, типа *золотой*, *янтарный* и т. п.), в текстах XIX в. – квазицвета (прилагательные типа *светлый*, *тёмный*, *бледный* и т. п.), а у поэтов XX в. – собственно прилагательные цвета. При ранжировании и анализе пятидесяти наиболее частотных прилагательных в каждом десятилетии выявлен существенный факт: в 28 случаях из 50 первые места занимают ахроматические цвета *чёрный* и *белый*.

Статья формулирует некоторые направления наших дальнейших исследований обозначения цвета в русском языке. Это прежде всего дальнейшая разработка классификации прилагательных цвета, статистические аспекты описания их полисемии и сопоставление частотности их употребления у отдельных авторов, что позволит в результате предложить целостный статистико-диахронический подход к описанию цветообозначений в русских поэтических текстах. Обозначены также некоторые другие важные направления, требующие историко-культурной и литературоведческой интерпретации статистической информации.

© Масевич А. Ц., Захаров В. П., 2021

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2021. Т. 19, № 4. С. 5–35
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2021, vol. 19, no. 4, pp. 5–35

Ключевые слова

цветообозначения, прилагательные цвета, диахронические исследования, статистика, поэтические тексты, частотность слов, поэтический корпус НКРЯ, семантическая разметка

Для цитирования

Масевич А. Ц., Захаров В. П. Диахронические изменения частотности цветообозначений в русских поэтических текстах // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2021. Т. 19, № 4. С. 5–35. DOI 10.25205/1818-7935-2021-19-4-5-35

Diachronic Changes in the Frequency of Color Designation in Russian Poetic Texts

Andrey Ts. Masevich¹, Victor P. Zakharov²

¹ St. Petersburg State Institute of Culture
St. Petersburg, Russian Federation

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

¹ andmasev@mail.ru, <https://orcid.org/0000-0002-1752-8915>

² v.zakharov@spbu.ru, <https://orcid.org/0000-0003-0522-7469>

Abstract

We propose a statistical-diachronic approach to the description of color designations in poetic texts in Russian. Methods of corpus linguistics and the semantic annotation of the Russian National Corpus (RNC) applied in the research have revealed certain patterns of change in the frequency of color designations in poetic texts from 1750 to 1999. A quantitative characteristic of the poetic corpus of the RNC is presented by decades of the period under study. We also provide a general description of its poetry corpus (number of texts, number of authors, searching techniques, the semantic annotation, etc.). Also presented is the authors' classification of adjectives with the color tag of the semantic annotation.

The study has revealed the RNC inconsistency in the assignment of color tags during the semantic annotation. In the authors' opinion, this defect could affect the results of the study; however, the spotted tendencies in frequency behavior of color adjectives seem to be reliable.

The research method is based on the comparison of statistical indicators for each decade in the specified range. Accordingly, the charts are plotted with reference to the average values for the range, unlike the RNC Charts tool, where a chart is based on the points for each year with a smoothing coefficient. Also, the indicators in the poetic corpus are compared with the corresponding data in the prose corpus. It has been reliably established that in the texts of the poetry corpus the frequency of color designations is higher than in the texts from basic, newspaper and oral corpora regardless of the part of speech the words belong to. The highest frequency (IPM) values of all adjectives that have the color tag of the RNC semantic annotation are observed in the last decades of the 19th - first decades of the 20th centuries. The frequency behavior of adjectives belonging to different classes differs significantly in different periods. In the texts of the 18th century, the analogous adjectives (or the adjectives of secondary nomination, such as *золотой* (golden), *янтарный* (amber), etc.) prevail. In the texts of the 19th century adjectives – quasi-colors, (such as *светлый* (light), *тёмный* (dark), *бледный* (pale), etc.) are most frequent whereas in the texts of the 20th century the color adjectives per se have the highest frequency. While ranking the data selected from all the periods in the descending order, the first 28 ranks were occupied by the adjectives denoting achromatic colors – black and white. The article also defines prospects for our further research of color designations in Russian: specifying classification of color adjectives, looking at statistical aspects of their polysemy, and comparison of their use frequency by individual authors.

Keywords

color designations, color adjectives, diachronic studies, statistics, poetic texts, word frequency, poetry corpus of the RNC, semantic annotation

For citation

Masevich, A. Ts., Zakharov, V. P. Diachronic Changes in the Frequency of Color Designation in Russian Poetic Texts. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2021, vol. 19, no. 4, pp. 5–35. (in Russ.) DOI 10.25205/1818-7935-2021-19-4-5-35

Введение

Перспективная цель настоящей статьи состоит в том, чтобы предложить схему статистико-диахронического описания цветообозначений в текстах различных типов на основе методов корпусной лингвистики.

Ранее нам уже удалось выявить некоторые важные закономерности в частотном поведении цветообозначений в русской поэзии [Захаров, Масевич, 2019; Масевич, Захаров, 2019а; 2019б; Masevich, Zakharov, 2019]. Было показано, что обозначение цвета в произведениях любого жанра реализуется в первую очередь за счет использования прилагательных. При этом частота встречаемости прилагательных цвета любого типа значительно выше в поэтических текстах, чем в текстах других жанров. Сравнение частотности прилагательных цвета по подкорпусам поэтических текстов разных авторов показывает большой разброс, причем в XX в. частотность прилагательных цвета выше, чем в другие периоды.

Разумеется, на данный момент нам удалось довольно подробно осветить лишь одну сторону этого явления, поскольку объект исследования ограничен поэзией и не включает тексты других функциональных типов. Выбор поэтических текстов обусловлен тем, что в них слова-цветообозначения в целом, в том числе и прилагательные цвета, употребленные в прямом или метафорическом значении, играют более значимую роль (для авторов иногда концептуальную), чем в каких-либо других литературных или речевых произведениях. Более подробно мы покажем это далее в данной статье, сравнив частоты цветообозначений в поэтических и прозаических текстах. При этом отметим, что частота нескольких отдельных прилагательных цвета в прозаических текстах художественной литературы несколько выше, чем в поэтических. Этот факт требует дополнительного исследования по конкретным прилагательным, в том числе с анализом контекстов употребления.

Частотное поведение прилагательных цвета рассматривается нами на значительном отрезке времени – с 1750 по 1999 г. В ходе исследования выявлен ряд статистико-диахронических фактов и закономерностей (паттернов), которые мы постарались представить как можно более полно. Вместе с тем их предметная интерпретация – литературоведческая, стилистическая, культурно-историческая – дело специалистов соответствующих областей, здесь мы ограничились осторожным предложением гипотез.

Поэтический корпус Национального корпуса русского языка (НКРЯ)

НКРЯ содержит 89 124 текста (документа), принадлежащих 951 автору с годами рождения с 1658 до 1967 г. (см. [Гришина и др., 2009]). В корпусе представлены тексты разных поэтических жанров, в том числе пьесы в стихах. Объем корпуса на 17 августа 2021 г. составляет 12 407 747 словоупотреблений.

Как видно на графиках (рис. 1, 2) в разрезе десятилетий наибольшие число текстов и их объем в корпусе относятся к 1910-м гг. (т. е. к Серебряному веку), при этом взрывной рост начинается в первой декаде XX в., а достигнутый уровень удерживается еще два-три десятилетия. Затем объем поэтического подкорпуса неуклонно снижается вплоть до конца 1990-х гг., опускаясь до уровня конца XVIII в. Еще один важный культурно-исторический факт: поэзия XX в. по числу текстов очень сильно превышает русскую поэзию XIX в.

Как видно из рис. 1 и 2, линейная зависимость между числом текстов и числом словоупотреблений отсутствует. Самыми «многословными» в истории русской поэзии оказались два десятилетия XX в. (1910–1919 и 1920–1929), затем идет культурно-исторический период первых лет советской власти, точнее, до 1959 г.

Выявленная динамика может быть обусловлена рядом внутри- и внелитературных факторов. Укажем на два из них – очевидных и взаимосвязанных: рост творческой активности в указанные периоды и / или расширение издательской деятельности в определенные исторические периоды. По-видимому, число словоупотреблений в поэтических текстах может свидетельствовать о творческой активности их авторов. Чем больше словоупотреблений, тем авторы активнее, тем больше создается поэтических произведений. Правда, нельзя утверждать, что суммарное число словоупотреблений указывает на качество или уровень поэзии периода. Произведения авторов Золотого века, безусловно, не ниже качеством, чем авторов Серебряного века.

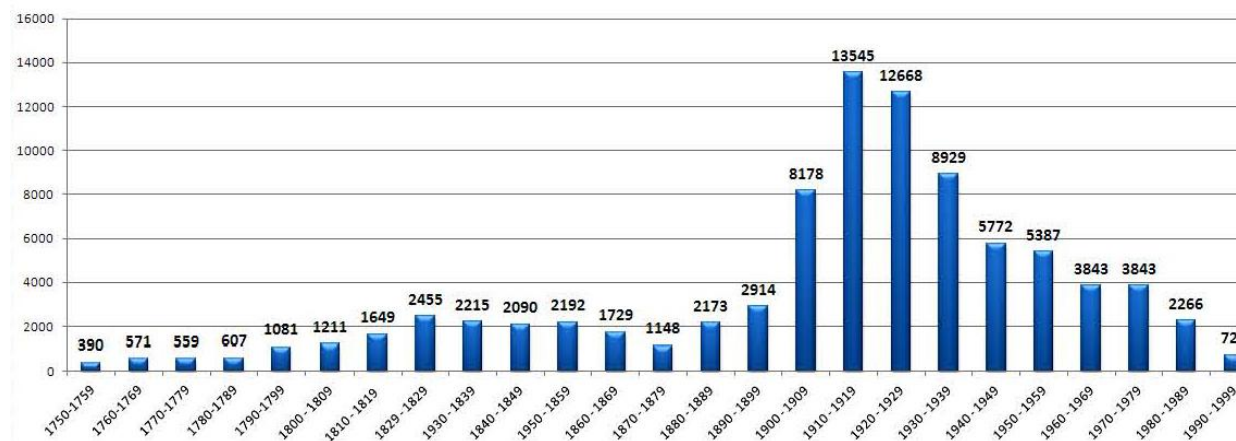


Рис. 1. Число текстов (документов) в ПК НКРЯ по десятилетиям с 1750 по 1999 г.
 Fig. 1. Number of texts (documents) in the RNC poetry corpus by decades from 1750 to 1999

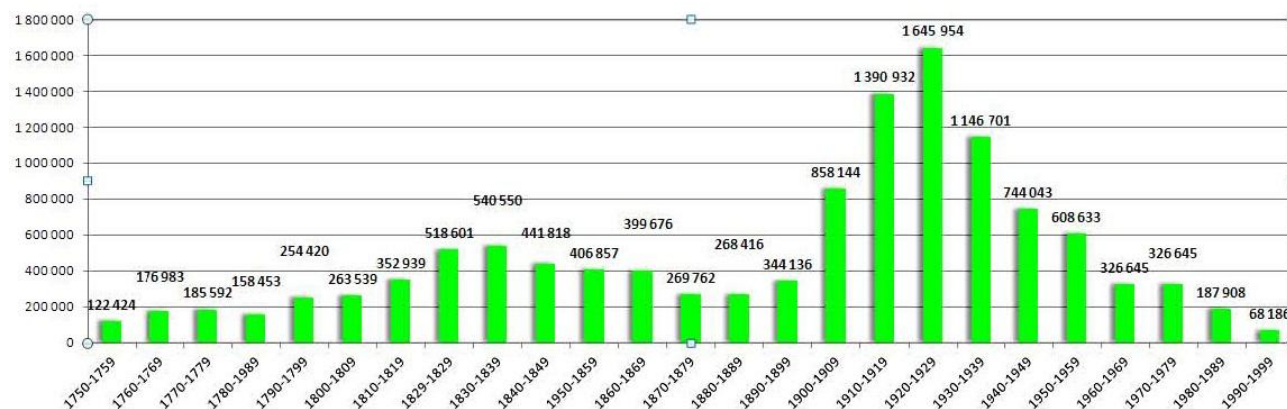


Рис. 2. Объемы подкорпусов (число словоупотреблений) в ПК НКРЯ по десятилетиям с 1750 по 1999 г.
 Fig. 2. Sizes of the subcorpora (number of token) in the RNC poetry corpus by decades from 1750 to 1999

Наша статья готовилась в основном на основе старой версии НКРЯ. Перепроверка данных показала, что в новой версии число словоупотреблений в поэтическом корпусе в девятые годы значительно выросло – 332 166 против 68 186 по старой версии (рис. 3), равно как и в предшествующие десятилетия.

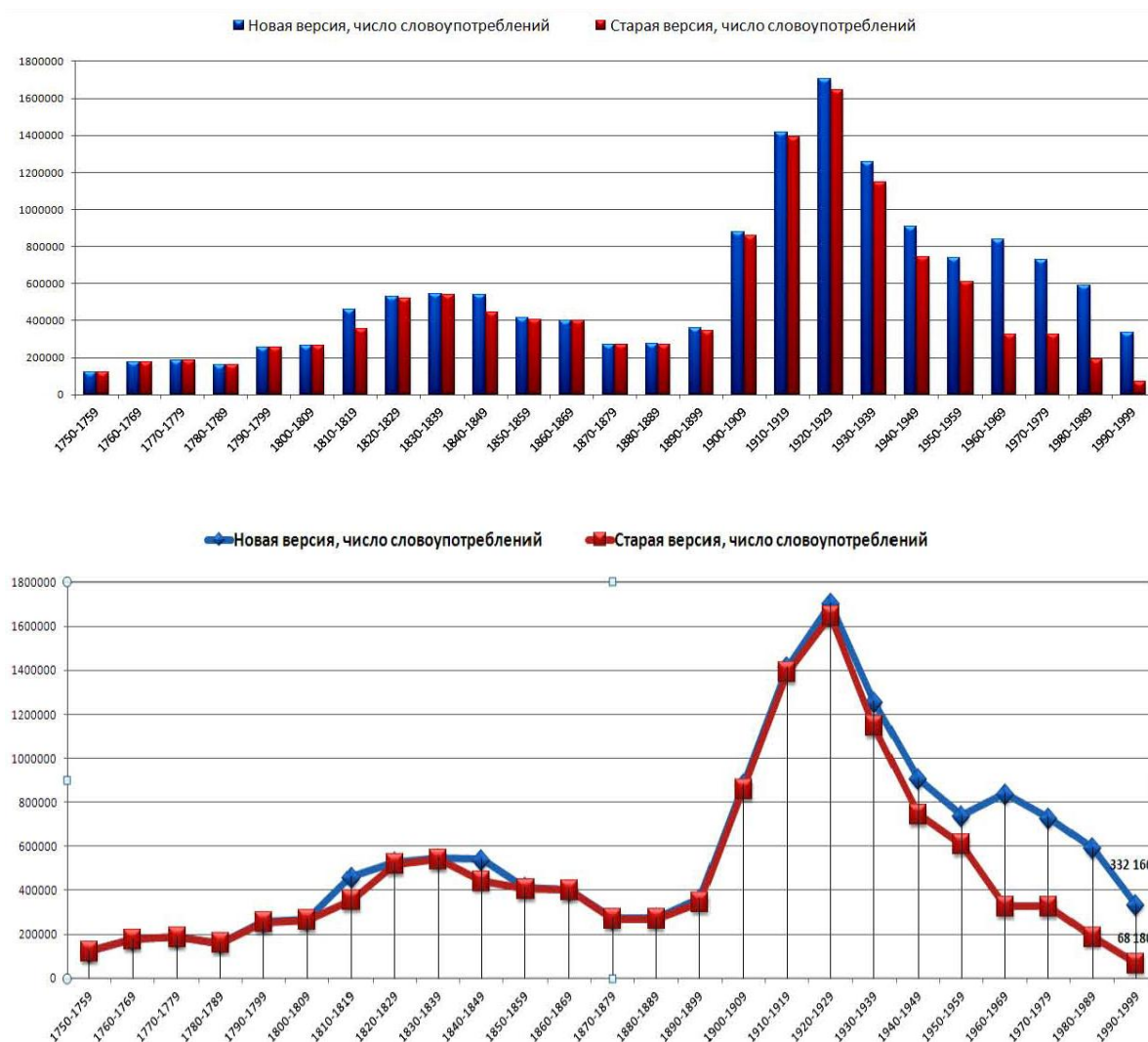


Рис. 3. Число словоупотреблений по декадам в старой и новой версиях ПК НКРЯ

Fig. 3. Number of tokens by decades in old and new versions of poetry corpus of RNC

Как видно на графиках, существенные различия между двумя версиями корпуса отмечаются только в последние четыре десятилетия XX в., что вполне логично – пополнение корпуса, по-видимому, осуществляется за счет более поздних изданий. Важно отметить, что в новой версии, в отличие от старой, отмечается рост числа словоупотреблений с 1960 по 1969 г.

Сопоставим наши подсчеты с данными других авторов [Lyashevskaya et al., 2018], представленными на рис. 4.

Сравним период 1960–1969 гг. с периодом 1950–1959 гг. (см. рис. 3) и период 1961–1970 гг. с периодом 1961–1970 гг. (см. рис. 4). Мы видим, что число словоупотреблений в поэтическом корпусе в 1960-е гг. выросло. Графики, представленные на рис. 3, с одной стороны, и на рис. 4, с другой, были построены независимо, различия состоят лишь в технике

и масштабах построения. Сопоставление собственно статистической информации показывает, что содержательно они практически полностью совпадают.

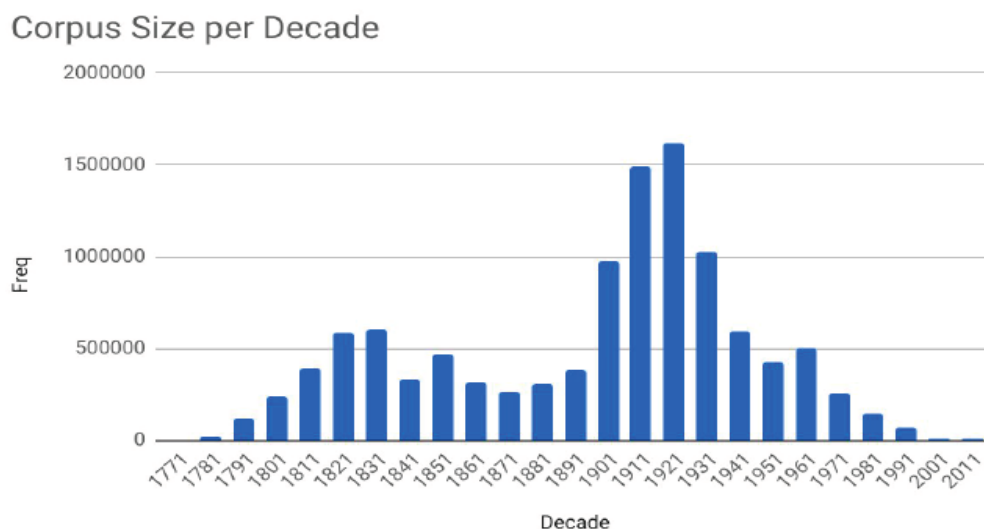


Рис. 4. Объемы подкорпусов (число словоупотреблений) в ПК НКРЯ с 1771 по 2011 г. по данным [Lyashevskaya et al., 2018]

Fig. 4. Sizes of the subcorpora (number of tokens) in the RNC poetry corpus by decades from 1771 to 2011 according to [Lyashevskaya et al., 2018]

При сравнении списков авторов старой и новой версий выявлено, что в старой версии отсутствовали такие имена, как Белла Ахмадулина, Андрей Вознесенский, Евгений Евтушенко, Сергей Стратановский, Елена Шварц и ряд других, определявших поэзию шестидесятых годов прошлого века.

Однако и в новой версии по-прежнему отсутствуют тексты многих значимых для русской поэзии второй половины XX в. авторов, таких как Анри Волохонский, Владимир Высоцкий, Михаил Генделев, Глеб Горбовский, Борис Гребенщиков, Юлия Друнина, Игорь Иртеньев, Тимур Кибиров, Лев Лосев, Новелла Матвеева, Булат Окуджава, Дмитрий Пригов, Роберт Рождественский, Вероника Тушнова, Алексей Хвостенко и ряд других, что несколько снижает репрезентативность поэтического корпуса НКРЯ и, следовательно, его надежность как исследовательского инструмента.

Сделаем предположение, что рост «поэтической активности» в 1960-е гг. связан с общественно-политической активностью в период оттепели. В остальное время в обеих версиях отмечается снижение этого показателя. Особо примечателен показатель 1990–1999 гг., когда объем поэтического корпуса опустился на уровень начала XIX в. (см. рис. 3).

Основываясь на приведенных данных, можно предположить, что к концу XX в. наметился «коллапс» русской поэзии. Это вполне согласуется с имеющимися в литературе утверждениями об умирании поэзии да и литературы вообще. Так, например, Д. Уолкот обнаружил этот феномен во французской культуре: «В то время, когда французская поэзия находится при последнем издыхании, они начинают торговать всякой тухлятиной французского литературоведения...» (цит. по: [Фефелов, 2015, с. 67]). Кроме того, общеизвестные статистические данные свидетельствуют о падении тиражей художественной литературы, о снижении читательского интереса.

«Коллапс» поэзии может иметь, в числе прочих, следующие признаки:

- 1) значительное снижение творческой активности и числа авторов;
- 2) сокращение издательской деятельности;

- 3) уменьшение интереса к поэзии;
- 4) снижение художественного уровня поэтических текстов.

Однако достаточно взглянуть на статистику портала Стихи.ру (рис. 5), чтобы подвергнуть сомнению наличие сегодня первого и отчасти второго признаков.

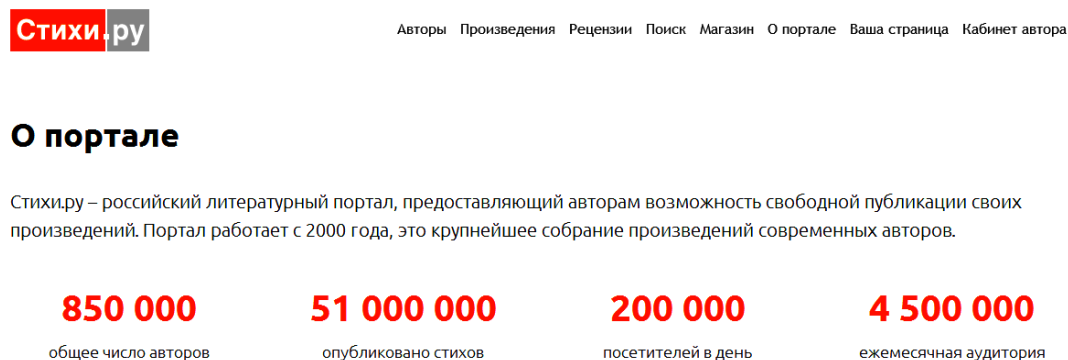


Рис. 5. Страничка портала Стихи.ру

Fig. 5. The page of the website stihhi.ru

Другой вопрос – как оценить объем аудитории? Еще труднее оценить художественные достоинства опубликованных произведений. С большой вероятностью, однако, можно предположить размещение значительного числа произведений невысокого качества.

Таким образом, наши данные позволяют говорить о «коллапсе» поэзии лишь предположительно. Нам кажется, что сегодня следует говорить скорее о радикальном изменении форм бытования поэзии в современной коммуникационно-технологической среде.

Семантическая разметка

В НКРЯ мы имеем первый (или один из первых) в истории корпусной лингвистики опыт семантической разметки национального корпуса [Куцова и др., 2005]¹. Семантическая разметка НКРЯ – это еще один важный фактор, который влияет на обеспечение поиска лексических единиц, имеющих общие семантические признаки.

Каждое словоупотребление (включая цветообозначения) имеет в корпусе полное описание, содержащее грамматические, семантические и некоторые другие признаки, представленные в виде кодов (тегов). Каждое описание имеет области: лемма, грамматика, семантика основная, семантика дополнительная и дополнительные признаки. На примере лексемы *красный* оно представлено на рис. 6.

Семантические признаки в поисковом запросе задаются с помощью вкладок меню (рис. 7), которые раскрываются по ссылке в запросной форме *семант. признаки: выбрать*. Каждое меню соответствует определенной части речи. Существительным соответствуют три меню – *предметные*, *непредметные* и *собственные имена*. К сожалению, в семантических признаках глаголов признак *цвет* отсутствует.

¹ Детальное описание разметки имеется на сайте НКРЯ (<https://ruscorpora.ru/new/corpora-sem.html>).

102. Б. А. Слудский. «Ткал ковры. И продавал — внарез...» (1970-1975) [омонимия не снята] [Все приме](#)

Нет, не должен. **Красное** и синее
вызывали чувства силь
Подходили!

[Б. А. Слудский. «Ткал ковры

103. Б. А. Слудский. Рассвет в музее

Как учат алфавиту: «А»
сначала альфу, а потом
твой **красный** полыхал
и твой зеленый обещал

[Б. А. Слудский. Рассвет в музее

Красное	
Лемма	красный (см. в словарях)
Грамматика	прил, с, ед, вин, полн, disamb прил, с, ед, им, полн, disamb
Семантика основная	r:qual, t:physq:color
Семантика дополнительная	der:shift, dt:physq:color, ev:posit, r:qual, r:rel, t:species
Доп. признаки	capital, first, словарн, numred, posred

[Сообщить об ошибке...](#)

Рис. 6. Описание словоформы *красное* в поэтическом корпусе НКРЯ

Fig. 6. Description of the word *красное* (red) in the RNC poetry corpus

www.ruscorpora.ru/reqsem.html

Имена: [предметные](#) | [непредметные](#) | [собственные](#) | [Прилагательные](#) | [Числительные](#) | [Местоимения](#) | [Глаголы](#) | [Наречия](#)

Прилагательные

Разряд

☐ качественные

☐ относительные

☐ приткательные

☐ неизменяемые

Такономия

☐ размер в том числе:

☐ большой ☐ малый

☐ абсолютный

☐ расстояние в том числе:

☐ большое ☐ малое

☐ количество в том числе:

☐ большое ☐ малое

☐ абсолютное

☐ место

☐ направление

☐ время в том числе:

☐ длительность в том числе:

☐ большая ☐ малая

☐ абсолютная

☐ возраст в том числе:

☐ большой ☐ малый

☐ абсолютный

☐ скорость в том числе:

☐ большая ☐ малая

☐ физические свойства в том числе:

☐ форма

☐ цвет

☐ вкус

☐ запах

☐ температура

☐ качества человека

☐ Оценка в том числе:

☐ положительная

☐ отрицательная

Словообразование

☐ диминутивы

☐ аугментативы

☐ аттенуативы

☐ хабитивы

☐ каритивы

☐ потенциальные в том числе:

☐ возможные

☐ невозможные

☐ отыменные прилагательные

в том числе образованные от имен со значением:

☐ лица

☐ животные

☐ растения

☐ вещества и материалы

☐ пространство и место

☐ здания и сооружения

☐ инструменты и приспособления

☐ еда и напитки

☐ тексты

☐ части тела и органы

☐ отлагательные прилагательные

☐ отадавербальные прилагательные

Рис. 7. Меню выбора семантических признаков для прилагательных


Fig. 7. The menu of semantic tags for adjectives

Результаты поиска сопровождаются статистическим описанием (рис. 8), в котором указывается объем поискового корпуса (подкорпуса), а именно число текстов (документов) и число словоупотреблений (слов), число текстов (документов), в которых найдено соответствие поисковому запросу, и число соответствий запросу (вхождений).

Вы находитесь в новой версии

Сообщить об ошибке

Подробнее



НАЦИОНАЛЬНЫЙ КОРПУС
РУССКОГО
ЯЗЫКА

Результаты поиска в поэтическом корпусе

[перейти на страницу поиска](#) [выбрать подкорпус](#) [версия с ударениями](#) [настройки](#) [формат KWIC](#) [English](#)

Объем всего корпуса: 89 124 документа, 12 407 747 слов.

красный

Найдено 4 601 документ, 6 240 вхождений.

[Распределение по годам](#)

Поискать в других корпусах: [основном](#), [акцентологическом](#), [газетном](#), [диалектном](#), [мультимедийном](#), [обучающем](#), [параллельном](#), [устном](#).

Рис. 8. Статистика результатов поиска по запросу *красный* в ПК НКРЯ
Fig. 8. Statistics of results of the search in the RNC poetry corpus

Методика исследования

Сначала для периода с 1750 по 1999 г. было создано 25 пользовательских поэтических подкорпусов, каждый из которых охватывал десятилетие. Средствами поиска по семантическому признаку из результатов поиска в каждом подкорпусе извлекался набор прилагательных цвета, после чего средствами MS Excel создавались таблицы для каждого периода и подсчитывалось значение IPM (instances per million) для каждой леммы. Далее, где необходимо, частные таблицы сливались в единую таблицу. Названия классов разработанной классификации (см. далее) вносились в таблицы вручную. Затем проводился анализ данных, и средствами MS Excel строились соответствующие графики.

Добавим к этому два существенных терминологических замечания. В нашем исследовании используются два базовых понятия: **обозначение цвета** и **частота**.

Для первого из них существует большое число альтернативных терминов, или квазитерминов: *цвет (цвета), оттенок цвета или цветов, обозначение цвета, цветообозначение* (как процесс и как номинация), *цветонаименование, колоративная лексика, колоратив, колороним, колоризм, имя цвета, цветковые прилагательные, цветные прилагательные* и т. п. Мы не будем вникать в сходства и различия между ними, но из стилистических соображений для понятия «цветообозначение» будем использовать следующие выражения, которые считаем синонимичными: *слова-цветообозначения; слова, обозначающие цвет; «цветные» слова / прилагательные, колоративы*.

Понятие «частота» может относиться к разным лингвистическим единицам, у нас это слова (иногда коллокации) или словоупотребления. В первом случае мы будем говорить о лексемах или леммах. Во втором используем термин «словоупотребление». Частота может быть абсолютная или относительная (нормированная), т. е. пересчитанная по отношению к объему корпуса или подкорпуса. Почти всегда наше исследование основывается на относительной частоте. Для обозначения этого понятия мы используем также сокращенный вариант «частота», иногда «частотность», а также единицу измерения, известную как IPM (ipm, instances per million). Значение IPM вычисляется как частное от деления абсолютной частоты на объем корпуса (выборки), умноженное на миллион.

Сопоставление частотности использования цветообозначений в поэтических и прозаических текстах

Цвета и цветové оттенки обозначаются знаменательными словами, относящимися к разным частям речи – существительным, наречиям и глаголам, но основную функцию описания цвета выполняют прилагательные (рис. 9).

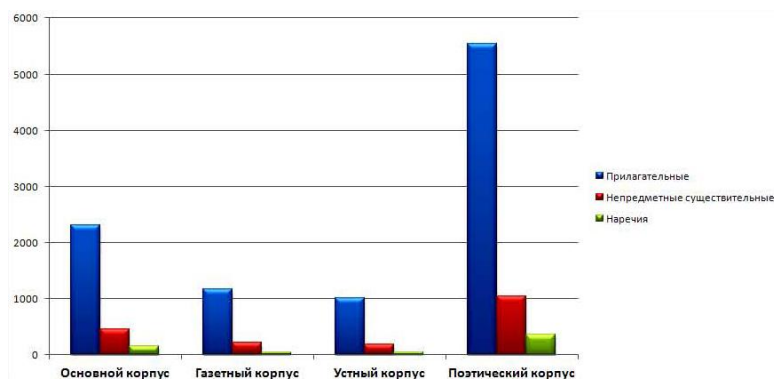


Рис. 9. Сопоставление относительной частоты (IPM) прилагательных, непредметных существительных и наречий, имеющих тег *цвет*, в четырех корпусах НКРЯ

Fig. 9. Comparison of IPM of adjectives in the Basic corpus, Newspaper corpus, Oral corpus and Poetry corpus

На рис. 9 видно, что: 1) ИРМ слов-цветообозначений, относящихся к прилагательным, выше, чем у существительных и наречий; 2) для всех трех частей речи ИРМ «цветных» слов в поэтическом корпусе значительно больше, чем в остальных трех корпусах.

Как уже говорилось, в НКРЯ у глаголов признак *цвет* отсутствует. Поэтому, чтобы подтвердить (или опровергнуть) вышеприведенные заключения для глаголов, мы провели дополнительное микроисследование, отобрав 13 наиболее характерных глаголов, содержащих в значении сему *цвет* и сравнив затем их частотные характеристики в тех же корпусах НКРЯ. Результаты поиска тринадцати глаголов-цветообозначений в четырех корпусах НКРЯ см. в табл. 1.

Таблица 1

Относительная частота глаголов, обозначающих цвет, в корпусах НКРЯ

Table 1

Relative frequency of verbs meaning colors in RNC corpora

Глагол (Verb)	ИРМ в корпусах НКРЯ (IPM in RNC corpora)			
	основной (Basic corpus)	газетный (Newspaper corpus)	устный (Oral corpus)	поэтический (Poetry corpus)
1) белеть	11,84	0,70	1,65	81,79
2) чернеть	10,47	0,59	0,50	66,84
3) краснеть	17,96	3,08	5,37	40,39
4) зеленеть	6,87	0,81	0,99	45,68
5) желтеть	4,77	0,59	0,99	28,72
6) синеть	6,78	0,78	0,25	63,01
7) сереть	2,00	0,27	0,08	11,22
8) розоветь	2,67	0,13	0,17	19,97
9) коричневеть	0,07	0,01	0,00	0,27
10) оранжеветь	0,01	0,00	0,00	0,09
11) буреть	0,52	0,18	0,25	0,64
12) алеть	2,64	0,23	0,91	30,73
13) лиловеть	0,35	0,00	0,00	4,10
Среднее значение ИРМ	5,15	0,57	0,86	30,27

Сопоставим теперь частотность 12-ти базовых прилагательных цвета в поэтических текстах (48 авторов) и в произведениях художественной прозы (18 авторов из основного корпуса НКРЯ) (рис. 10).

Отметим, что у трех из двенадцати прилагательных цвета *красный*, *серый* и *коричневый* ИРМ в текстах художественной прозы несколько выше, чем в поэтических. В остальных девяти случаях выше частотность прилагательных в поэтических текстах, при этом, однако, разница у прилагательных *жёлтый* и *розовый* очень небольшая, а у прилагательных *белый* и *чёрный* не столь значительна, как у *синий*, *голубой*, *зелёный*.

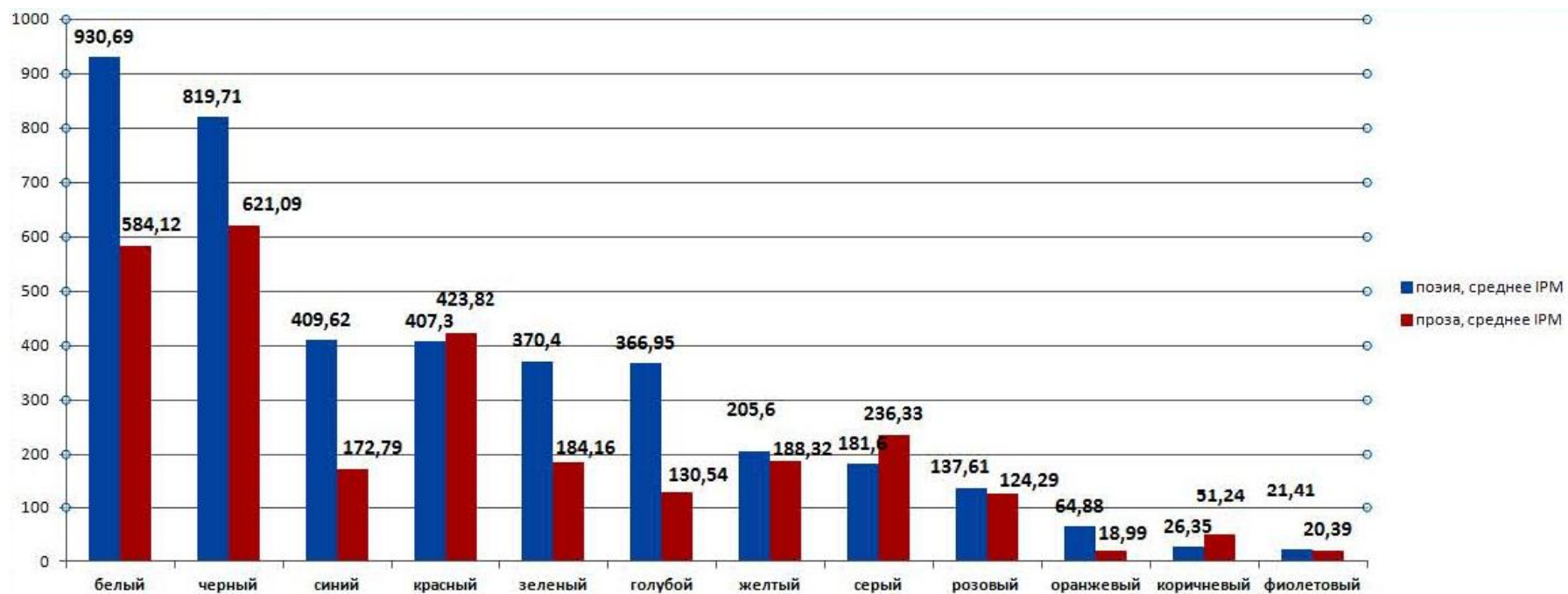


Рис. 10. Сравнение средней частоты двенадцати прилагательных цвета в поэтических текстах и в художественной прозе (сортировка по убыванию IPM)

Fig. 10. Comparison of average values of frequency of color adjectives in the poetry texts and in the fiction texts

Закljučая этот раздел, зафиксируем следующие наблюдения. При сравнении числа цветообозначений по частям речи и типам текстов частота «цветных» прилагательных для всех типов текста выше, чем частота «цветных» слов других частей речи. Среднее ИРМ тринадцати глаголов, обозначающих цвет, в текстах поэтического корпуса также значительно выше, чем в текстах основного, газетного и устного корпусов (см. табл. 1).

ИРМ слов цветообозначений (т. е. имеющих тег *цвет* в НКРЯ) – прилагательных, существительных и наречий – в текстах поэтического корпуса значительно выше, чем в текстах основного, газетного и устного корпусов (см. рис. 9).

ИРМ двенадцати прилагательных цвета (отобраны слова, обозначающие основные цвета или базовые цветовые категории) в поэтических текстах 48 авторов за тремя исключениями (*красный, серый, коричневый*) выше, чем в текстах 18 авторов художественной прозы (см. рис. 10). Однако это превышение не столь выражено, как при сравнении частотности цветообозначений в поэтическом корпусе с частотностью в целом в основном, газетном и устном корпусах (см. рис. 9). Наши данные показывают близость лексики поэзии и лексики художественной прозы. Эти факты, однако, требуют дополнительного, уже содержательного, исследования контекстов употребления и, возможно, более представительных подкорпусов. Если говорить конкретнее, то предметом таких исследований может быть связь частотного поведения слов-цветообозначений и функционального стиля прозаических текстов.

Классификация прилагательных, имеющих тег семантической разметки *цвет* в НКРЯ

Отличие предлагаемой нами классификации от других состоит в том, что она составлена *a posteriori* и относится к прилагательным, имеющим тег *цвет* в НКРЯ, иначе говоря, к материалу, с которым мы работаем.

Всего в поэтическом корпусе НКРЯ 1 641 454 словоупотребления прилагательных, в том числе 102 817 словоупотреблений прилагательных, имеющих тег *цвет*, что составляет 6,3 % от всех словоупотреблений прилагательных.

Предлагаемая нами схема классификации, содержит семь классов².

1. Прилагательные собственно цвета.
2. Прилагательные квазицвета (*светлый, тёмный, бледный, пёстрый* и т. п.). Слова этой группы не означают цвета, а только характеризуют его, поэтому Р. М. Фрумкина [1984], например, не относит их к именам цвета. Тем не менее, значительная часть этих слов имеет в НКРЯ тег *цвет*, и поэтому они (не без оснований) включены в нашу классификацию.
3. Аналоговые прилагательные цвета (слова вторичной номинации), например *малиновый, вишнёвый, янтарный, золотой* и т. п.
4. Объектные прилагательные ограниченной сочетаемости (например, *гнедой, саврасый, пегий* и т. п.).
5. Диминутивы (*беленький, синенький* и т. п.)
6. Прилагательные подобию (неполноты качества): *розоватый, зеленоватый* и т. п.
7. Сложные (составные) прилагательные: *тёмно-синий, нежно-голубой* и т. п.

Приведенная классификация нам необходима, чтобы сопоставить частотное поведение различных групп прилагательных, семантически относящихся к цвету. На ее основе мы создавали и дифференцировали массивы лексики для диахронических исследований.

Эта классификация была описана нами ранее [Захаров, Масевич, 2019; Масевич, Захаров, 2019б], однако позднее мы сформулировали некоторые дополнительные основания и классификационные признаки, а также детализировали ряд классов. Новое описание классифика-

² Отметим, что позиции 5–7 могут рассматриваться как подклассы класса «дериваты». Однако в нашем исследовании при построении графиков такая двухуровневая детализация осложнила бы методику, не давая при этом какой-либо существенной информации.

ции будет дано в отдельной статье, здесь же мы используем уже введенные ранее классы, что достаточно для достоверного показа качественного изменения лексики русских поэтических текстов на протяжении трех веков.

Семантическая разметка НКРЯ, которая является одним из наиболее важных инструментов нашего исследования, имеет ряд существенных недостатков, в частности, непоследовательность присвоения тегов. Так, некоторые прилагательные, отнесенные нами к квазицветам (например, *светлый, тёмный, пёстрый, бледный, клетчатый* и др.), имеют тег *цвет*, а другие (например, *матовый, яркий, жухлый, радужный* и др.) его не имеют.

Такая же непоследовательность отмечается в разметке прилагательных вторичной номинации, которые мы назвали аналоговыми. Так, прилагательные *аметистовый, опаловый, дымчатый, малахитовый* имеют тег *цвет*, а *вишнёвый, малиновый, лимонный, горчичный* и др. этого тега не имеют.

Еще сложнее положение с составными прилагательными. В одних случаях система разметки НКРЯ рассматривает такое прилагательное как одну лексическую единицу, например *тёмно-синий*, и присваивает ему тег *цвет*, в других, например *ярко-красный*, трактует его как две отдельные лексические единицы.

Более или менее последовательно размечены лишь прилагательные собственно цвета.

Непоследовательность разметки, конечно же, в некоторой степени искажает приводимые здесь количественные показатели, но всё-таки не влияет на общую картину изменения частотности обозначений цвета в диахронии.

Количественные характеристики прилагательных цвета во временных периодах

Для исследования в поэтическом корпусе были созданы подкорпусы по датам написания текстов, причем каждый из подкорпусов охватывает десятилетие за период 1750 по 1999 г.

Число «цветных» лексем в поэтическом корпусе по десятилетиям в XIX в. (рис. 11) плавно растет и достигает некоторого пика в 1860-е гг., затем несколько снижется в 1870-е, и после этого снова отмечается рост с пиком уже в двадцатые годы XX в., после чего мы видим плавное снижение до конца заданного периода.

Отметим, что приведенные на рис. 1, 2, 12 графики отражают динамику абсолютных значений связанных между собой величин. Рассмотрим теперь относительные частоты (IPM) (рис. 13).

Интересно, что пик значений IPM на рис. 13 приходится на 1900-е гг., в то время как на других графиках (см. рис. 11, 12) он приходится на 1910-е или 1920-е гг. Это значит, что «цветная» лексика этого десятилетия отличается наибольшим богатством и разнообразием.

При сопоставлении динамики частот прилагательных цвета в поэтическом корпусе и в основном корпусе (рис. 14), мы видим:

- 1) во все временные периоды частота прилагательных, имеющих тег *цвет*, в поэтическом корпусе выше, чем в основном;
- 2) тенденции диахронического «поведения» прилагательных с тегом *цвет* в поэтическом и основном корпусах приблизительно сходны;
- 3) изменения в основном корпусе выражены значительно слабее, чем в поэтическом.

Сравнение частоты классов прилагательных с тегом *цвет* по десятилетиям с 1750 по 1999 г. (рис. 15) показывает следующее:

- 1) частотность прилагательных собственно цвета, квазицветов и аналоговых значительно выше, чем у объектных прилагательных (т. е. прилагательных ограниченной сочетаемости), диминутивов, прилагательных подобия (неполноты качества) и составных прилагательных;
- 2) аналоговые прилагательные более частотны в текстах XVIII в.;
- 3) квазицвета более частотны на протяжении всего XIX в.;
- 4) в XX в. частота прилагательных собственно цвета значительно выше, чем у других классов прилагательных.

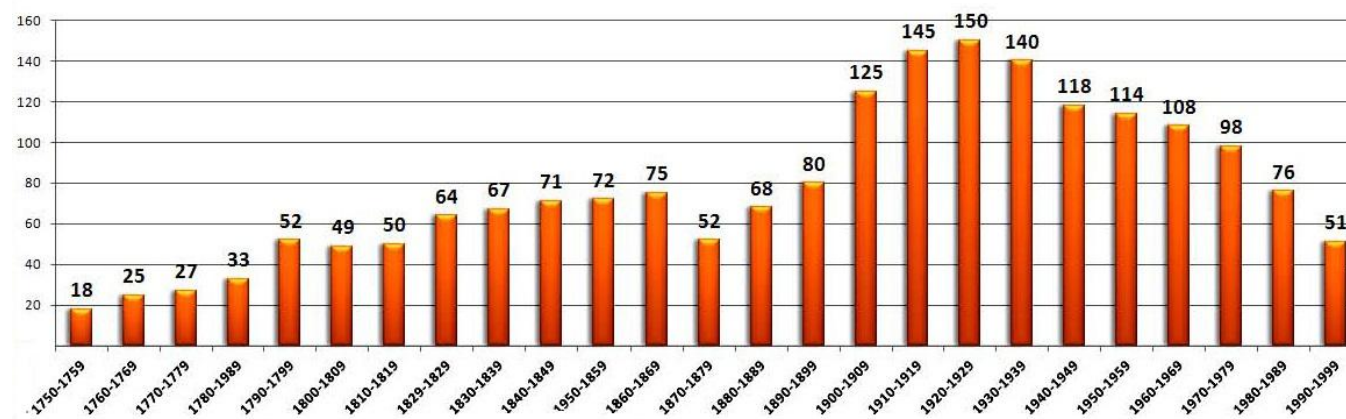


Рис. 11. Число лексем прилагательных с тегом *цвет* в текстах ПК НКРЯ по десятилетиям с 1750 по 1999 г.
 Fig. 11. Number of adjective lexemes with *color* tag in the texts of the RNC poetry corpus by decades from 1750 to 1999

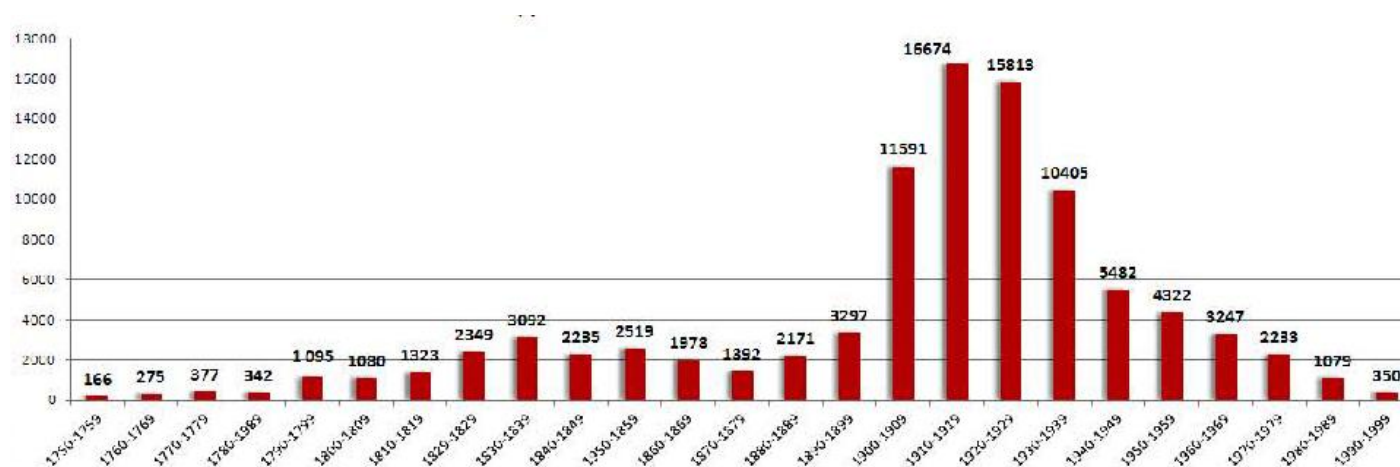


Рис. 12. Число словоупотреблений прилагательных с тегом *цвет* в подкорпусах НКРЯ по десятилетиям с 1750 по 1999 г.
 Fig. 12. Number of word usages with *color* tag in the texts of the RNC poetry corpus by decades from 1750 to 1999



Рис. 13. ИРМ прилагательных с тегом *цвет* в ПК НКРЯ по десятилетиям с 1750 по 1999 г.
 Fig. 13. IPM of adjectives with *color* tag in the texts of the RNC poetry corpus by decades from 1750 to 1999

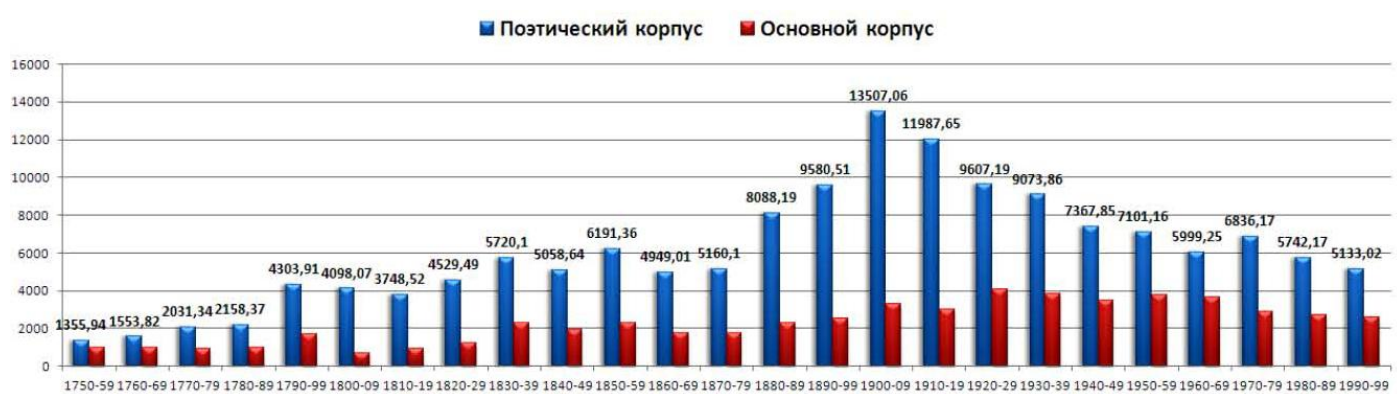


Рис. 14. Диахроническое сравнение частоты прилагательных с тегом *цвет* по десятилетиям в поэтическом и основном корпусах НКРЯ
 Fig. 14. Diachronic comparison of frequency values of adjectives with *color* tag in the RNC poetry corpus and the RNC basic corpus

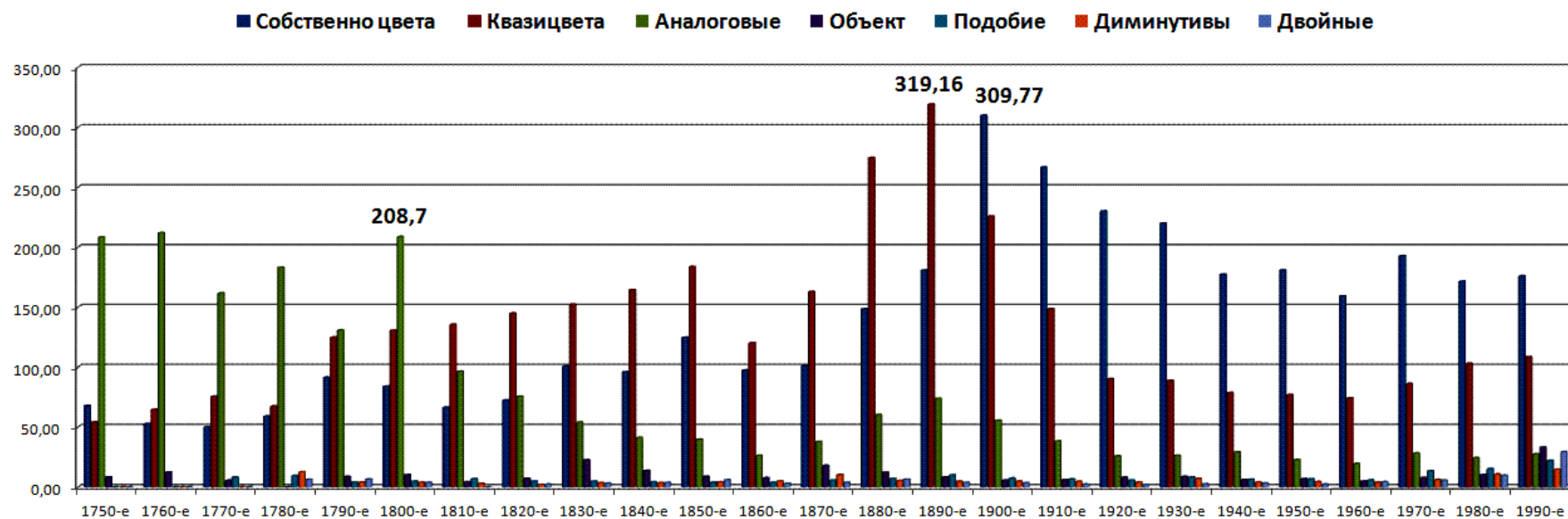


Рис. 15. Диахроническое сравнение среднего IPM классов прилагательных с тегом *цвет* по десятилетиям с 1750 по 1999 г.
 Fig. 15. Diachronic comparison of average IPM values of adjectives with color tags by decades from 1750 to 1999

При анализе графика на рис. 15 следует помнить отмеченные выше недостатки семантической разметки. Тем не менее, мы считаем, особенно учитывая надежность единиц измерения – относительной частоты, что наши данные отражают реальную динамику. Эта динамика, конечно, нуждается в объяснении, но на данный момент мы его не имеем, и это, видимо, задача будущих исследований.

Построим диаграммы, позволяющие увидеть частотное поведение различных прилагательных цвета отдельно по каждому веку (рис. 16–18).

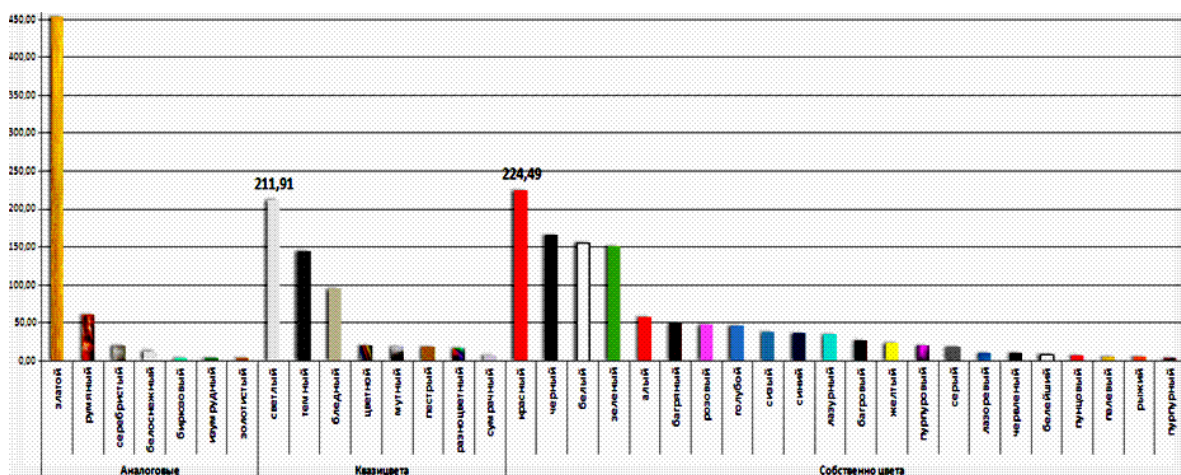


Рис. 16. IPM прилагательных с тегом *цвет* в поэтическом корпусе НКРЯ за период 1750–1799 гг.

Fig. 16. IPM of adjectives with color tag in the RNC poetry corpus within period 1750–1799

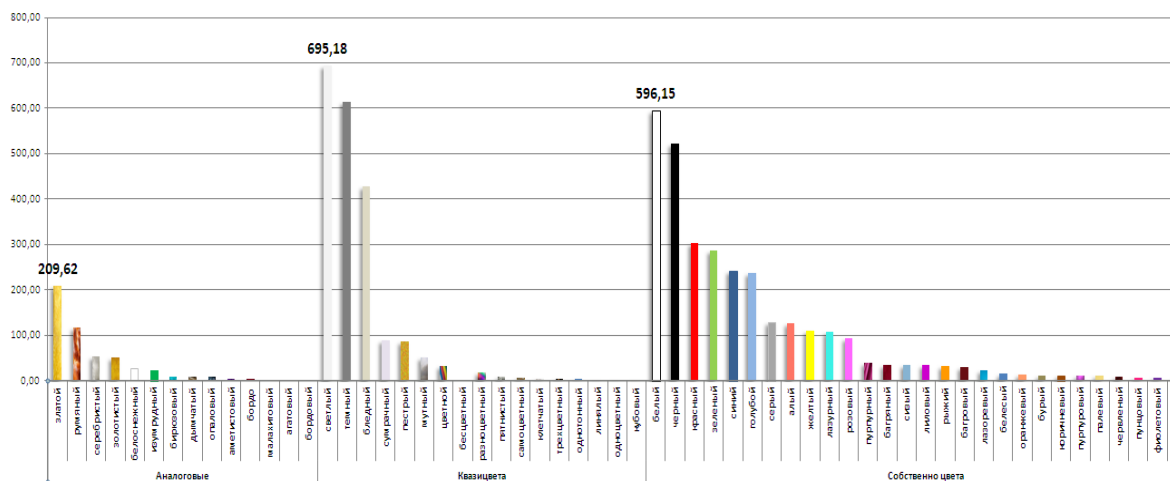


Рис. 17. IPM прилагательных с тегом *цвет* в поэтическом корпусе НКРЯ, за период 1800–1899 гг.

Fig. 17. IPM of adjectives with color tag in the RNC poetry corpus within period 1800–1899

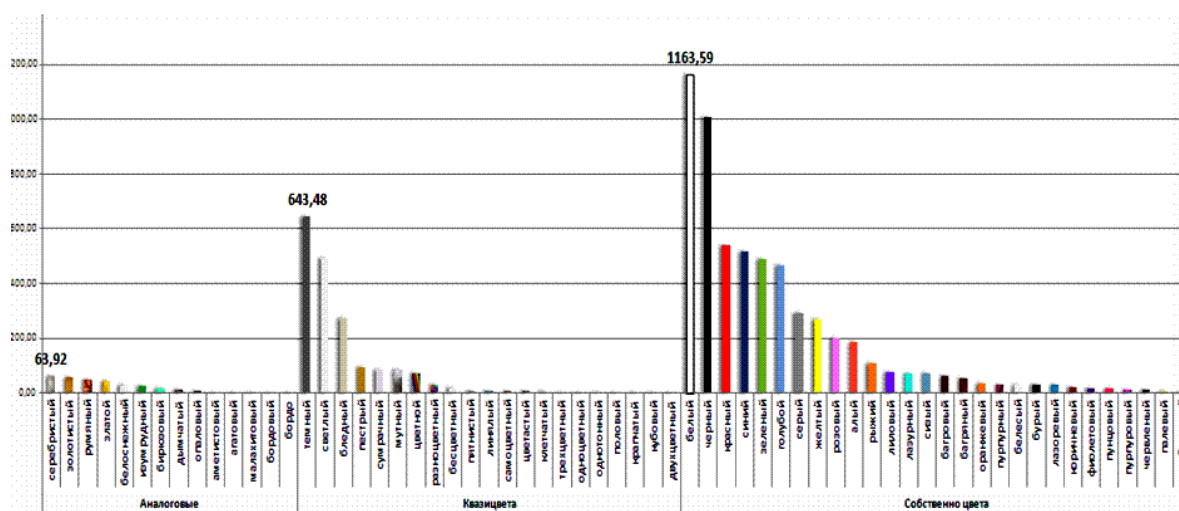
Рис. 18. IPM прилагательных с тегом *цвет* в поэтическом корпусе НКРЯ за период 1900–1999 гг.

Fig. 18. IPM of adjectives with color tag in the RNC poetry corpus within period 1900–1999

На рис. 16 видно, что высокая частотность аналоговых прилагательных обусловлена главным образом частотностью лексемы *золотой*. По-видимому, дело в многозначности этой лексемы. Помимо значения цвета, она, как это свойственно аналоговым прилагательным, означает *сделанный из золота*, а также имеет очень распространенное метафорическое значение *счастливый*. В XIX в., как видно на рис. 17, более высокий IPM отмечается у квазицветов, причем наиболее частотными оказываются три лексемы этого класса – *светлый*, *тёмный* и *бледный*. В XX в. наиболее высокий IPM у лексем *белый*, *чёрный*, означающих ахроматические цвета (см. рис. 18). Отметим также, что «лидеры» среди квазицветов лексемы *тёмный* и *светлый* по значению IPM выше или близки к лексемам *красный*, *синий*, *зелёный*, *голубой*.

Анализ частотного поведения четырех частотных прилагательных

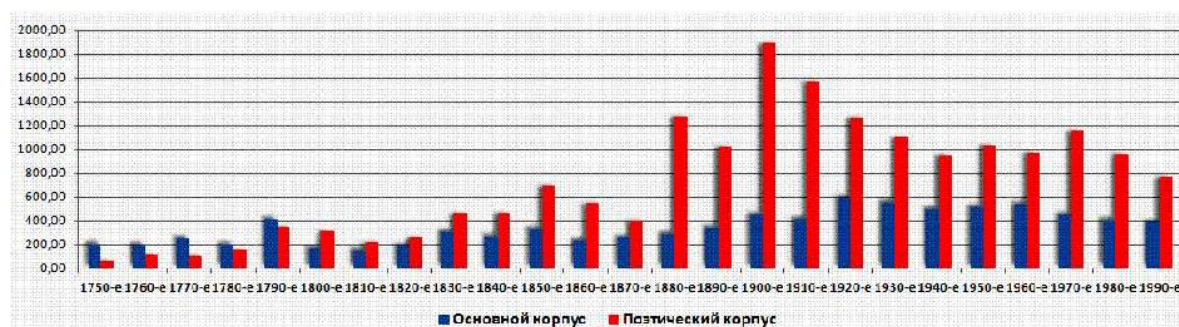
Сравним далее диахроническую динамику пар ахроматических прилагательных цвета и соответствующих им квазицветов (*белый* – *светлый* и *чёрный* – *тёмный*) в поэтическом и основном корпусах (рис. 19, 20).

В XVIII в. IPM прилагательного *белый* в основном корпусе выше, чем в поэтическом (см. рис. 19, а). В последующие периоды наблюдается обратное соотношение. При этом разница между IPM в основном и в поэтическом корпусах в XX в. выражена сильнее, чем в предыдущие периоды. Можно также отметить, что временные колебания значений IPM в текстах поэтического корпуса выражены значительно сильнее.

В отличие от прилагательного *белый* прилагательное *светлый* встречается чаще, и намного, в поэтическом корпусе на протяжении всего заданного периода (см. рис. 19, б).

Поведение лексемы *чёрный* (см. рис. 20, а) в значительной степени сходно с поведением лексемы *белый* как в основном, так и в поэтическом корпусе (см. рис. 19, а).

Поведение прилагательного *тёмный* (см. рис. 20, б), как можно наблюдать, сходно с поведением прилагательного *светлый*.

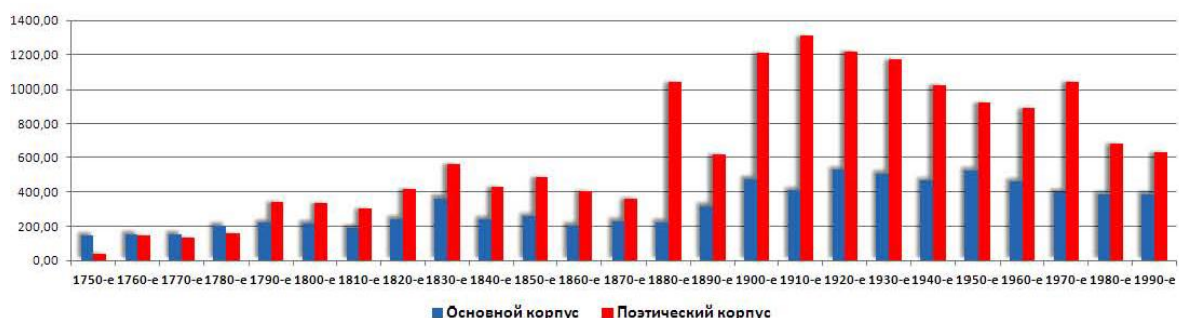


а



б

Рис. 19. Изменение значений ИРМ прилагательных *белый* (а) и *светлый* (б) по десятилетиям в поэтическом и основном корпусах НКРЯ
 Fig. 19. Changing of values of IPM of adjectives *белый* 'white' (a) and *светлый* 'light' (b) by decades in the RNC poetry corpus and RNC basic corpus



а



б

Рис. 20. Изменение значений ИРМ прилагательных *чёрный* (а) и *тёмный* (б) по десятилетиям в поэтическом и основном корпусах НКРЯ
 Fig. 20. Changing of values of IPM of adjectives *чёрный* 'black' (a) and *тёмный* 'dark' (b) by decades in the RNC poetry corpus and RNC basic corpus

Далее сравним средние значения частотности тех же четырех прилагательных в поэтическом и основном корпусах НКРЯ по векам написания текстов (рис. 21).

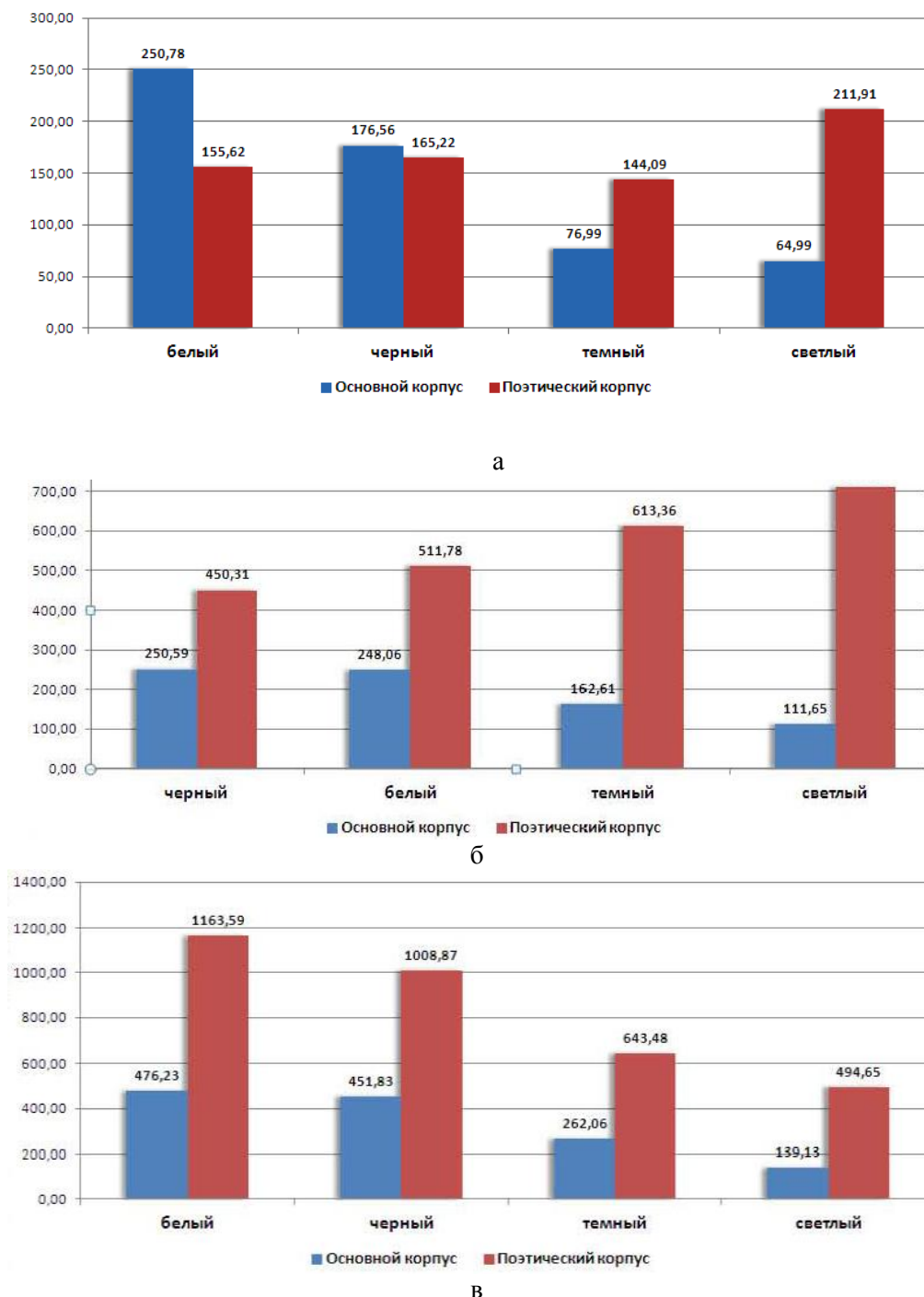


Рис. 21. Сравнение среднего значения ИРМ прилагательных *белый*, *чёрный*, *тёмный* и *светлый*:
 а – в текстах XVIII в. (с 1750 по 1799 г.); б – в текстах XIX в. (с 1800 по 1899 г.);
 в – в текстах XX в. (с 1900 по 1999 г.)

Fig. 21. Comparison of the average value of IPM of adjectives *белый* 'white', *чёрный* 'black', *тёмный* 'dark' and *светлый* 'light': а – in texts of 18th century (from 1750 to 1799);
 б – in texts of 19th century (from 1800 to 1899); в – in texts of 20th century (from 1900 to 1999)

На рис. 21, *а* отчетливо видны особенности частотного поведения четырех прилагательных в текстах второй половины XVIII в. ИРМ прилагательных *чёрный* и *белый*, т. е. прилагательных собственно цвета, в основном корпусе выше, чем в поэтическом, а ИРМ квазицветов *светлый* и *тёмный* выше в поэтическом корпусе.

В XIX в. (рис. 21, *б*) ИРМ прилагательных *белый*, *чёрный*, *тёмный*, *светлый* в поэтическом корпусе выше, чем в основном. При этом в поэтических текстах их частота друг относительно друга (в данном порядке) растет, в то время как в основном корпусе НКРЯ она уменьшается. Иначе говоря, в поэтическом корпусе мы видим более высокую частотность квазицветов, а в основном – выше частотность прилагательных собственно цвета.

В XX в. ИРМ всех четырех слов в поэтическом корпусе выше, чем в основном. В отличие от XIX в. в обоих корпусах частотность прилагательных собственно цвета выше частотности квазицветов. Выскажем здесь осторожную гипотезу: по-видимому, это один из признаков сближения языка поэзии и языка прозы в XX в.

Наиболее частотные прилагательные цвета в поэтических текстах

Таблицы частотности всех прилагательных по десятилетиям были слиты в единую таблицу за весь рассматриваемый период (ранг – лексема – ИРМ – десятилетие, всего 2 004 строки), которая была отсортирована по убыванию ИРМ. Естественно, одни и те же лексемы в сводной таблице повторяются, будучи представлены в разных десятилетиях с разной частотой.

Рассмотрим первые 50 строк с рангами от 1 до 50. Нетрудно догадаться, что это будут, как правило, одни и те же прилагательные, в первую очередь четыре вышеприведенные. Но интересно посмотреть на этот список по десятилетиям: какое из них представлено бóльшим числом разных прилагательных цвета с наиболее высокими значениями ИРМ. Сортировка таблицы по параметру «десятилетие» показывает, что это снова 1900-е гг. (см. рис. 13): из десяти высокочастотных прилагательных с рангом от 1 до 50 в этом десятилетии представлены девять лексем (*белый*, *бледный*, *голубой*, *зелёный*, *красный*, *светлый*, *синий*, *тёмный*, *чёрный*). В целом картина по десятилетиям в данном аспекте представлена в табл. 2. Эта таблица еще раз указывает на особый характер поэтического языка конца XIX – начала XX в.

Таблица 2

Число лексем с наиболее высоким ИРМ
по десятилетиям

Table 2

Number of lexemes with the highest IPM
within decades

Десятилетия (Decades)	Число лексем (Number of lexemes)	Десятилетия (Decades)	Число лексем (Number of lexemes)
1900-е гг.	9	1940-е гг.	2
1910-е гг.	6	1950-е гг.	2
1880-е гг.	4	1960-е гг.	2
1890-е гг.	4	1970-е гг.	2
1920-е гг.	4	1990-е гг.	2
1980-е гг.	4	1790-е гг.	1
1850-е гг.	3	1830-е гг.	1
1930-е гг.	3	1840-е гг.	1
Всего лексем 50			

Наконец, покажем 25 самых частотных прилагательных, в том числе находящихся за пределами ранга 50 в сводной таблице всех прилагательных, имеющих в поэтическом корпусе НКРЯ семантический тег *цвет* и встретившихся в нем хотя бы один раз (табл. 3). Отсортируем их в порядке возрастания рангов. Все эти слова, кроме прилагательных *златой* (1790-е гг.) и *пёстрый* (1850-е гг.), а также *рыжий* (1920-е гг.), имеют наибольшую частоту в конце XIX – начале XX в. (см. столбец «Десятилетие»).

Таблица 3

Прилагательные цвета с наиболее высокими значениями IPM
в поэтическом корпусе НКРЯ

Table 3

Adjectives of color with the highest values of IPM
in the RNC poetic corpus

Лексема (Lexeme)	Ранг первого появления лексемы в сводной таблице (The range of the first appearance)	Десятилетие (Decade)
1. белый	1	1900-е гг.
2. темный	3	1900-е гг.
3. черный	4	1910-е гг.
4. светлый	12	1890-е гг.
5. бледный	13	1890-е гг.
6. красный	29	1900-е гг.
7. синий	31	1900-е гг.
8. зеленый	34	1900-е гг.
9. златой	39	1790-е гг.
10. голубой	43	1900-е гг.
11. алый	68	1900-е гг.
12. серый	98	1910-е гг.
13. желтый	125	1900-е гг.
14. розовый	161	1900-е гг.
15. лазурный	181	1880-е гг.
16. румяный	210	1880-е гг.
17. сумрачный	224	1880-е гг.
18. рыжий	231	1920-е гг.
19. пестрый	238	1850-е гг.
20. золотистый	242	1900-е гг.
21. лиловый	275	1900-е гг.
22. сизый	276	1900-е гг.
23. серебристый	279	1900-е гг.
24. мутный	287	1910-е гг.
25. багряный	300	1900-е гг.

Можно с высокой степенью вероятности предположить, что в табл. 3 собраны 25 «любимых» цветов русской поэзии. Обратим внимание на то, что в списке нет нескольких прилагательных, означающих базовые цвета – *коричневый*, *оранжевый* и *фиолетовый*.

Золотой век и Серебряный век русской поэзии в зеркале статистики прилагательных цвета

В заключение сравним использование прилагательных цвета на протяжении двух периодов, которые в литературе получили название «Золотой век» и «Серебряный век» русской поэзии. Эти образные названия периодов в истории русской поэзии (и, шире, литературы) не имеют четко определенных хронологических рамок, но вместе с авторами статьи «Инструмент анализа данных в корпусе русской поэзии» [Lyashevskaya et al., 2018]³ мы будем относить Золотой век к 1811–1841 гг. и Серебряный век к 1901–1917 гг.

В указанной статье сравнивается частота прилагательных в поэтических текстах в двух исследуемых периодах, в частности приводится таблица с прилагательными, относительная частота которых в Серебряном веке больше, чем в Золотом веке. В этой таблице 54 прилагательных, из них 14 являются прилагательными цвета. Приведем эти прилагательные по статье О. Ляшевской и соавторов [Lyashevskaya et al., 2018] (табл. 4).

Таблица 4

Сравнение частоты прилагательных цвета
в поэтических текстах Золотого и Серебряного века

Table 4

Comparison of the frequency of adjectives of color
in poetic texts of the Golden and Silver ages

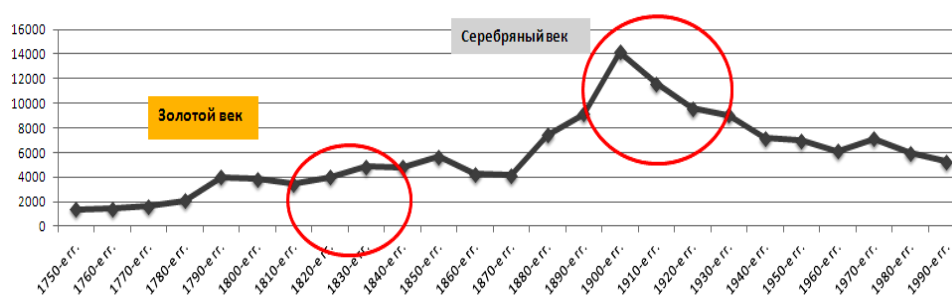
Прилагательное (Adjective)	Частота (Frequency)		F2/F1
	F1 (1811–1840)	F2 (1901–1917)	
серый	0,000615	0,003634	5,9
алый	0,000717	0,003747	5,2
розовый	0,000549	0,002124	3,9
белый	0,004053	0,014510	3,6
синий	0,001749	0,005925	3,4
тусклый	0,000388	0,001264	3,3
зелёный	0,001983	0,005704	2,9
красный	0,002275	0,006181	2,7
прозрачный	0,000885	0,002400	2,7
голубой	0,001836	0,004647	2,5
пёстрый	0,000571	0,001308	2,1
чёрный	0,004990	0,010419	2,1
тёмный	0,004968	0,010297	2,1
золотой	0,003702	0,007400	2,0

Из табл. 4 можно вывести, что частота прилагательных цвета в поэтических текстах первых двух десятилетий XX в. выше в среднем в 3,17 раза, чем в первой половине XIX в. В целом это соответствует нашим данным. Если посчитать число прилагательных цвета в указанных периодах (см. рис. 12), объем корпусов (см. рис. 2) и, соответственно, ИРМ, то, по нашим данным, $F2/F1 = 2,71$ ($12\ 568 / 4\ 645$). Это соотношение указывает, во сколько раз частота прилагательных цвета в поэтических текстах первых двух десятилетий XX в. выше соответствующего показателя за первые 40 лет XIX в. Небольшое расхождение между данными статьи О. Н. Ляшевской и нашими допустимо и объяснимо.

³ Статья О. Ляшевской с соавторами опубликована на английском языке, мы цитируем ее по-русски в нашем переводе.

Теперь сопоставим графики, построенные по нашей методике (назовем ее методика М1), и графики, построенные с помощью функции НКРЯ *Графики* (методика М2). Наша методика предполагает следующий порядок действий: поиск в подкорпусах ПК НКРЯ, сведение результатов поиска в электронной таблице и построение диаграмм по отдельным прилагательным по десятилетиям. Различия обусловлены тем, что система НКРЯ строит график по отдельным годам со степенью сглаживания⁴ (в данном случае она равна 8), и каждому году соответствует точка графика (М2). У нас же графики строились по декадам, и каждой декаде на оси ординат соответствует среднее значение частоты всех прилагательных цвета (или конкретного прилагательного) в этом десятилетии (М1).

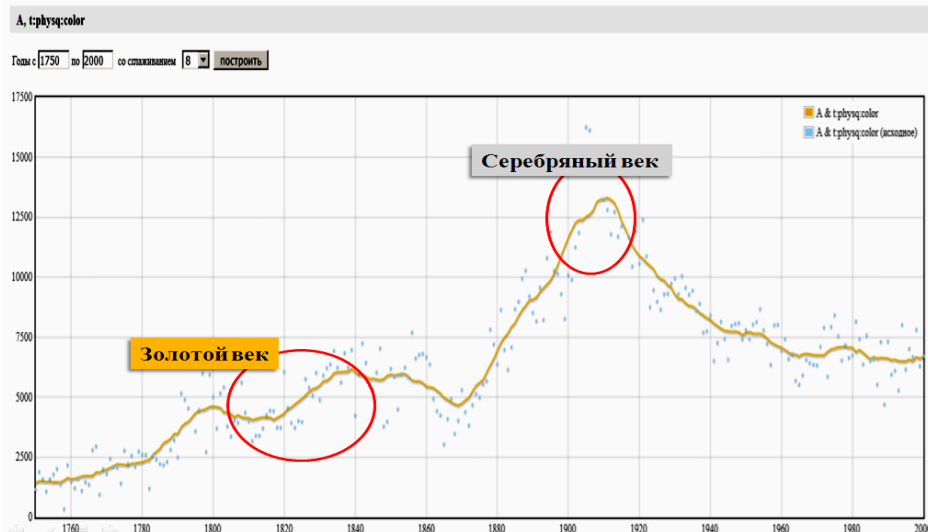
Покажем графики изменения частоты употребления во времени всех прилагательных цвета в соответствии с М1 (рис. 22, а) и М2 (рис. 22, б), а также для некоторых конкретных прилагательных (рис. 23–25). На всех графиках красными кружками обозначены сегменты, соответствующие Золотому веку и Серебряному веку.



а

Распределение по годам (частота на миллион словоформ) с 1750 по 2000

[перейти на страницу поиска](#)



б

Рис. 22. Значения IPM прилагательных, имеющих тег *цвет*, в текстах ПК НКРЯ.

Графики построены по М1 (а) и М2 (б)

Fig. 22. Values of IPM of adjectives with color tag in texts of the RNC poetry corpus.

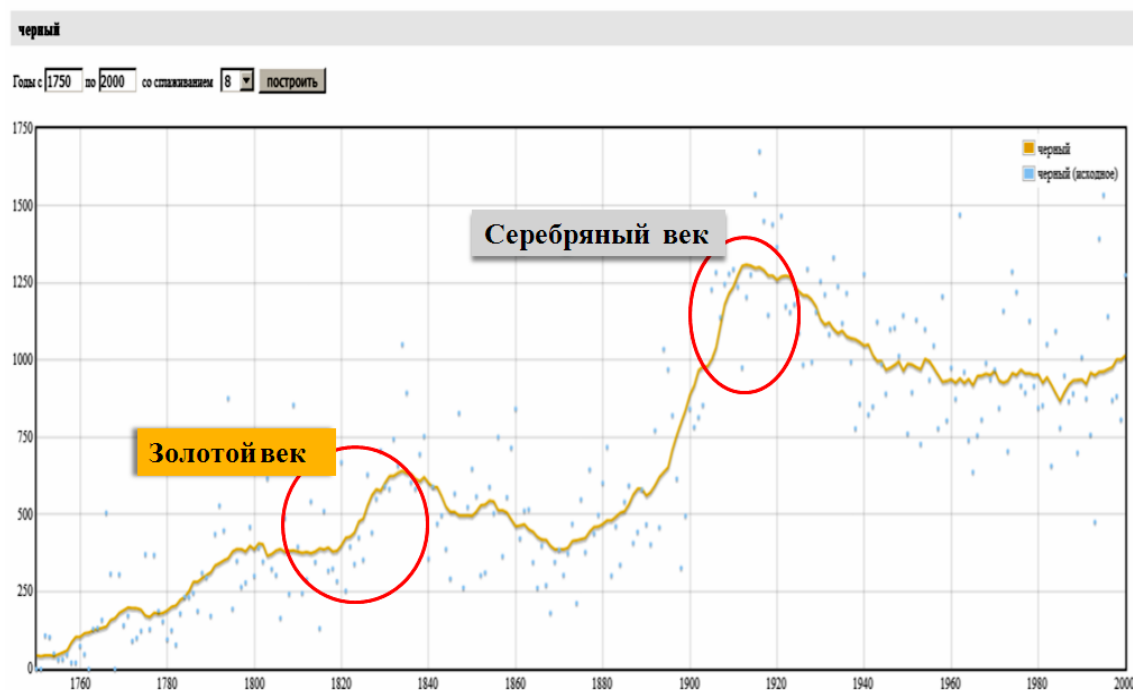
The charts are built according to М1 (а) and М2 (б)

⁴ «Сглаживание графиков позволяет увидеть общую тенденцию за случайными колебаниями частот. Например, сглаживание в 10 лет усредняет частоту слова с учетом предшествующих и последующих 5 лет» (<https://ruscorpora.ru/new/graphic.html>).



а

Распределение по годам (частота на миллион словоформ) с 1750 по 2000

[перейти на страницу поиска](#)

б

Рис. 23. Изменение значений IPM прилагательного *чёрный* в текстах ПК НКРЯ.

Графики построены по М1 (а) и М2 (б)

Fig. 23. Changing of values of IPM of adjective *чёрный* 'black' in the RNC poetry corpus.

The charts are built according to M1 (a) and M2 (b)

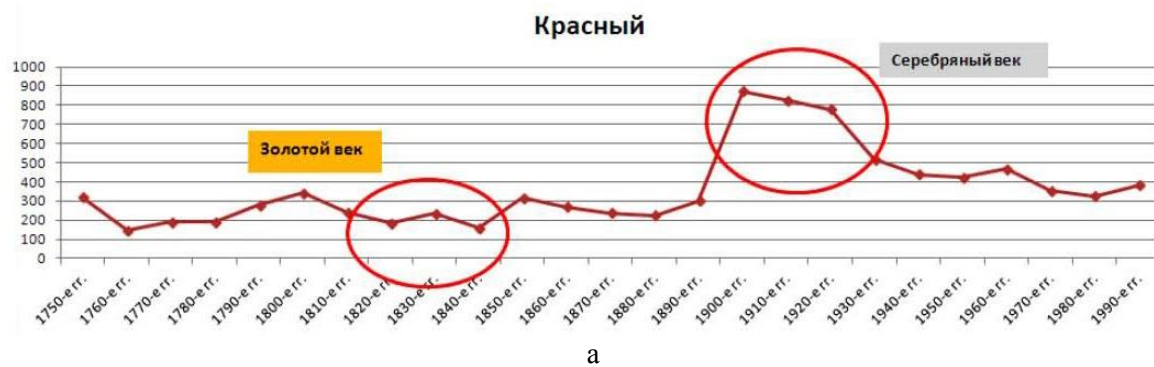
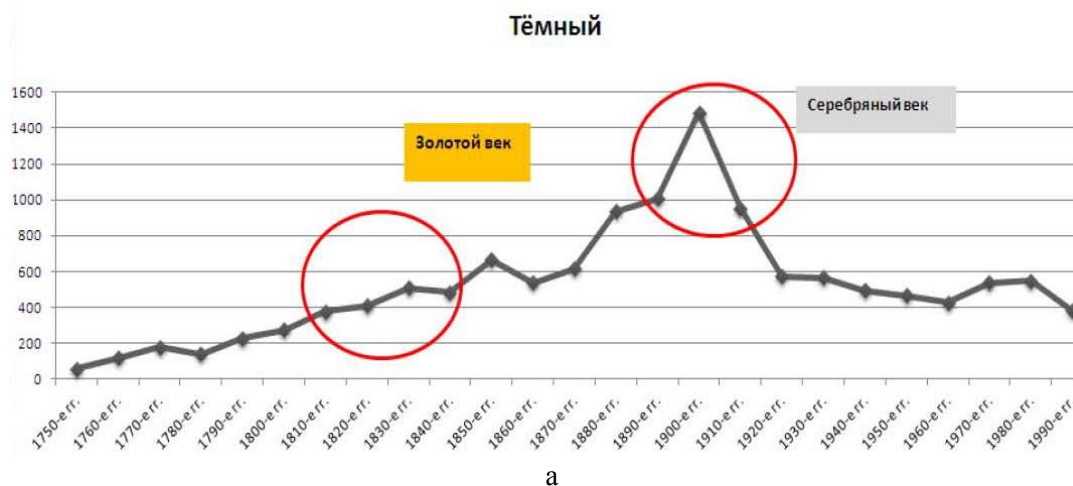


Рис. 24. График изменяемости IPM прилагательного *красный* в текстах ПК НКРЯ.

Графики построены по M1 (а) и M2 (б)

Fig. 24. Changing of values of IPM of adjective *красный* 'red' in the RNC poetry corpus.

The charts are built according to M1 (a) and M2 (b)



а



б

Рис. 25. График изменяемости ИРМ прилагательного *тёмный* в текстах ПК НКРЯ.

Графики построены по M1 (а) и M2 (б)

Fig. 25. Changing of values of IPM of adjective *тёмный* 'dark' in the the RNC poetry corpus.

The charts are built according to M1 (a) and M2 (b)

Мы наблюдаем, что при некоторых различиях в конфигурации кривых, построенных по разным методикам, графики изоморфны (небольшие отличия вряд ли существенны). Этот факт является дополнительным подтверждением достоверности наших данных. Обращает на себя внимание то, что в Золотом веке практически не наблюдается выраженных пиков частотности, тогда как в период Серебряного века пики частотности всех отобранных прилагательных цвета проявляются очень четко.

Выводы и заключение

В ходе исследования было обнаружено, что семантическая разметка слов-цветообозначений в НКРЯ в ряде случаев страдает неполнотой и непоследовательностью. В одном случае это нарушение логики разметки позволило нам выделить в его текстах целый класс специфичных прилагательных, которые мы назвали «квазицвета». В него входят прилагательные, относящиеся к цвету, но не обозначающие его. Этот класс позволил выявить существенные особенности функционирования цветообозначений.

Тем не менее, используя существующую семантическую разметку НКРЯ, мы выявили паттерны частотного поведения прилагательных цвета в русских поэтических текстах за период с 1750 по 1999 г. Сделаем некоторые выводы на их основе.

- В поэтических текстах наблюдается более высокая частотность словоупотреблений – обозначений цвета, чем в других типах текстов.

- При сравнении числа цветообозначений по частям речи и типам текстов частота цветообозначений-прилагательных во всех типах текстов выше частоты других частей речи.

- Графический паттерн частотного поведения определен как для всей совокупности прилагательных цвета (точнее, прилагательных, имеющих тег *цвет* в НКРЯ), так и для некоторых отдельных прилагательных.

- Выявленные нами тенденции частотного поведения совпадают с данными, представленными другими авторами.

- Самые высокие значения частоты всех прилагательных цвета мы видим в последнее десятилетие XIX – первые три десятилетия XX в.

- Для этого же периода характерно расширение словаря прилагательных цвета всех классов.

- Частотное поведение прилагательных, относящихся к разным классам, существенно отличается в разные временные периоды. С 1750 по 1809 г. преобладают аналоговые прилагательные (т. е. прилагательные вторичной номинации). Это обусловлено чрезвычайно высокой относительной частотой прилагательного *златой*, которое, по сути, маркирует всю эту эпоху.

- На протяжении практически всего XIX в., с 1810 по 1899 г., частота у прилагательных квазицвета выше, чем у прилагательных других классов.

- Начиная с первых лет XX в. и до его конца частота прилагательных собственно цвета (цвета как такового) значительно выше, чем у прилагательных других классов. При этом необходимо отметить, что частота прилагательных квазицвета *тёмный* и *светлый* в XX в. остается достаточно высокой.

- Наиболее частотными в нашем материале являются прилагательные, означающие ахроматические цвета *белый*, *чёрный* и квазицвета *светлый*, *тёмный* и *бледный*.

- Выявлены 25 наиболее частотных прилагательных цвета за весь исследуемый период («любимые» цвета русской поэзии). Их наибольшая концентрация пришлась на первые десять лет XX в. (1900–1909), что, видимо, не случайно. Это десятилетие находится в середине периода очень высокой творческой активности в русской поэзии, так называемой эпохи Серебряного века.

Динамика частотности прилагательных цвета определяется множеством факторов, влияющих на развитие языка. Выявление этих факторов и, соответственно, интерпретация наших данных требует отдельных историко-культурных и литературоведческих исследований. В одной из предыдущих публикаций [Масевич, Захаров 2019а, с. 24–28] мы отметили некоторую корреляцию частотного поведения прилагательных цвета с эволюционной моделью Берлина – Кея [Berlin, Kay, 1969; Kay, McDaniel, 1978]. Теперь предстоит понять причины изменения частотности прилагательных цвета в русской поэзии.

Хочется надеяться, что наши подходы и результаты, прошлые и настоящие, добавят что-то новое в представления о цветообозначении в русском языке. Кажется бесспорным, что

использование лингвистических корпусов открывает новые горизонты в исследовании цветообозначений и других лингвистических явлений.

В частности интерес могут представлять статистико-диахронические исследования различных значений прилагательных цвета. Собственно говоря, мы уже выполнили такое исследование о прилагательном *красный* (в печати). В настоящее время в работе находится исследование об особенностях использования прилагательных цвета различными авторами поэтических текстов. Интересной, как указано выше, представляется перспектива исследования взаимосвязи между частотностью цветообозначений и функциональными стилями. Особенно привлекательным нам видится возможное использование наших данных и методик в историко-культурных и литературоведческих исследованиях. В числе прочего это обеспечивало бы сопоставимость получаемых результатов.

При этом необходимо учитывать, что создание корпусов осуществляется преимущественно в автоматическом режиме, и по этой причине вряд ли можно всегда обеспечить полную надежность данных. Из этого следует, что полученные результаты должны проверяться всеми возможными способами.

Список литературы

- 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.
- Захаров В. П., Масевич А. Ц. Категоризация прилагательных цвета в русских поэтических текстах (корпусное исследование) // Тр. Ин-та рус. яз. им. В. В. Виноградова РАН. 2019. № 21. С. 164–181.
- 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.: Индрик, 2005. С. 155–174.
- Масевич А. Ц., Захаров В. П. Частотное поведение прилагательных цвета в русских поэтических текстах // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019а. Т. 17, № 1. С. 21–48. DOI 10.25205/1818-7935-2021-17-1-21-48
- Масевич А. Ц., Захаров В. П. Семантическое поле обозначений цвета в русском языке // Компьютерная лингвистика и вычислительные онтологии: Тр. XXII Междунар. объединенной науч. конф. «Интернет и современное общество», IMS-2019, Санкт-Петербург, 19–22 июня 2019 г. СПб.: ИТМО, 2019б. Вып. 3. С. 61–76.
- Фефелов А. Ф. Современное российское переводоведение: в поисках новой суверенной парадигмы // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2015. Т. 13, № 1. С. 48–72.
- Фрумкина Р. М. Цвет, смысл, сходство: Аспекты психолингвистического анализа. М.: Наука, 1984. 175 с.
- Berlin, B., Kay, P. Basic Color Terms: Their Universality and Evolution. Berkeley, University of California Press, 1969, 178 p.
- Kay, P., McDaniel, Ch. K. The Linguistic Significance of Meanings of Basic Color Terms. *Language*, 1978, no. 54 (3), pp. 610–646.
- Lyashevskaya, O., Litvintseva, K., Vlasova, E., Sechina, E. A Data Analysis Tool for the Corpus of Russian Poetry. In: HSE Working papers WP BRP 77/LNG/2018, Moscow, National Research University Higher School of Economics Press, 2018. URL: <https://ideas.repec.org/p/hig/wpaper/77-lng-2018.html> (accessed 04.09.2021).
- Masevich, A., Zakharov, V. Quantitative Analyses of Using Adjectives of Color in Russian Poetic Texts. In: CEUR Workshop Proceedings, 2019, vol. 2552, pp. 121–139. (in Russ.) URL: <http://ceur-ws.org/Vol-2552/Paper11.pdf> (accessed 04.09.2021).

References

- Berlin, B., Kay, P.** Basic Color Terms: Their Universality and Evolution. Berkeley, University of California Press, 1969, 178 p.
- Fefelov, A. F.** Paradigmatic Trends in the Evolution of the Russian Translation Studies: Pro and Contra Roman Jakobson. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2015, vol. 13, no. 1, pp. 48–72. (in Russ.)
- Frumkina, R. M.** Color, Meaning, Similarity: Aspects of Psycholinguistic Analysis. Moscow, Nauka, 1984, 175 p. (in Russ.)
- 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: National Corpus of the Russian Language: 2006–2008. New Results and Prospects. St. Petersburg, Nestor-History, 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.
- 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: National Corpus of the Russian Language: 2003–2005. Results and Prospects. Moscow, Indrik, 2005, pp. 155–174. (in Russ.)
- Lyashevskaya, O., Litvintseva, K., Vlasova, E., Sechina, E.** A Data Analysis Tool for the Corpus of Russian Poetry. In: HSE Working papers WP BRP 77/LNG/2018, Moscow, National Research University Higher School of Economics Press, 2018. URL: <https://ideas.repec.org/p/hig/wpaper/77-lng-2018.html> (accessed 04.09.2021).
- Masevich, A. Ts., Zakharov, V. P.** Semantic Field of Color Designations in Russian. In: Computational Linguistics and Computational Ontologies: Proc. of the XXII International Joint Scientific Conference “Internet and Modern Society”, IMS-2019, St. Petersburg, June 19–22, 2019) St. Petersburg, ITMO Uni. Press, 2019, iss. 3, pp. 61–76. (in Russ.)
- Masevich, A., Zakharov, V.** Frequency Behavior of Color Adjectives in Russian Poetic Texts. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, pp. 21–48. (in Russ.) DOI 10.25205/1818-7935-2021-17-1-21-48
- Masevich, A., Zakharov, V.** Quantitative Analyses of Using Adjectives of Color in Russian Poetic Texts. In: CEUR Workshop Proceedings, 2019, vol. 2552, pp. 121–139. (in Russ.) URL: <http://ceur-ws.org/Vol-2552/Paper11.pdf> (accessed 04.09.2021).
- Zakharov V. P., Masevich A. Ts.** Categorization of Adjectives of Color in Russian Poetic Texts (Corpus Research). In: Proceedings of the V. V. Vinogradov Institute of Russian Language. Russian Academy of Sciences, 2019, no. 21, pp. 164–181. (in Russ.)

Информация об авторах

Масевич Андрей Цезаревич, старший преподаватель

Захаров Виктор Павлович, кандидат филологических наук, доцент

Information about the Authors

Andrey Ts. Masevich, Lecturer

Victor P. Zakharov, Candidate of Sciences (Philology), Associate Professor

Статья поступила в редакцию 07.05.2021;
одобрена после рецензирования 01.07.2021; принята к публикации 01.07.2021
The article was submitted 07.05.2021;
approved after reviewing 01.07.2021; accepted for publication 01.07.2021