

Quantitative Analyses of Using Adjectives of Color in Russian Poetic Texts*

Andrey Masevich¹

andmasev@mail.ru

Viktor Zakharov²

v.zakharov@spbu.ru

¹ Saint-Petersburg State Institute of Culture,

² Saint Petersburg State University,
Saint Petersburg, Russia

Abstract

The paper deals with frequency behavior of adjectives of color, which was researched on the base of semantic annotation of Russian National Corpus (RNC) and standard computer technologies (MS Excel). The authors suggest a classification scheme for color adjectives tagged with semantic markup of RNC. The obtained results show the model of frequency behavior of color adjectives in poetic texts within a period from 1750 to 1999. It was revealed that IPM of color adjectives grows significantly at the beginning of 20th century. The frequency behavior of these adjectives correlates to their classes.

Keywords: *Russian National Corpus, semantic annotation, adjectives of color, classification, frequency behavior, poetic texts, diachronic studies*

1 Introduction

The quantitative characteristics of linguistics models and the notable frequency behavior of words give an idea of the real functioning of these models and add as much as possible to completeness, clarity and certainty.

The key research tool is the semantic annotation of the Russian National Corpus (RNC). However, it should be borne in mind that the semantic annotation of the RNC, implemented automatically, is still imperfect. Below we will show what its imperfection lies in. This raises the question of how the imperfection of annotation affects the results and, accordingly, how it should be taken into account in their evaluation and interpretation. We do not have an exact answer, although there is some indirect evidence in favor of the reliability of our results.

*Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

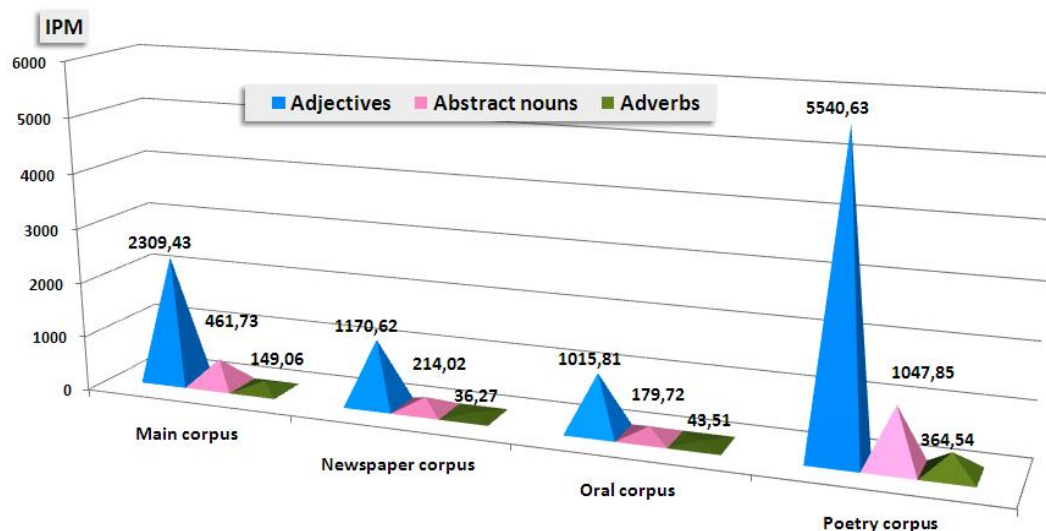


Figure 1: Comparison of IPM of adjectives, abstract nouns and adverbs denoting color in the four corpora RNC

The paper considers the frequency behavior, i.e. the change in the frequency of the use of adjectives of color (coloratives) in Russian poetic texts over a significant period of time – from 1750 to 1999. The work continues our previous studies [Masevich, Zakharov, 2019].

In addition to presenting the results of the study, the article demonstrates some possibilities of corpus linguistics and offers a methodology of semantic research using linguistic corpora and standard computer technologies.

2 Material Under Study

In this paper we consider quantitative characteristics of adjectives of color (i.e. adjectives denoting color) in Russian poetic texts. Our research is based on the material of the RNC corpora, mainly on the poetry corpus (PC RNC).

The choice for the study of poetic texts and adjectives is due to the fact that it is in them that the designation of color, direct or metaphorical, plays a more significant role than in any other literary or speech works, which is reflected in the frequency behavior of words of different lexical groups.

Colors are described by different content words – nouns, adverbs, and verbs, but the main function of describing color is performed by adjectives. Let us present in the form of a histogram the results of the research into the four corpora of the RNC (Fig. 1).

From Fig. 1 it is clear that the relative frequency (IPM) of color words related to adjectives, nouns and adverbs in the poetic corpus is much higher than in the other three corpora. It is also clearly seen that the IPM of adjectives is higher than in other parts of speech.

Unfortunately, in the RNC search menu for the verbs such characteristics as “color” is missing. Therefore, we selected 13 of the most characteristic verbs that contain the meaning of color, and compared their frequencies in the same four cases of RNC (Table 1).

From these data, it can be seen that the frequency of words meaning colors (or words

with the tag “color”) in the poetic corpus is also higher than their frequency in other corpora.

Thus, we can definitely state that the frequency of the terms of color in poetic texts is higher than in any genre of prose.

3 Semantic Annotation of RNC

The RNC demonstrates the first (or at least one of the first) experience of semantic annotation of the national corpus in the short history of corpus linguistics [Kustova et al., 2005]. Detailed description of the marking is available on the RNC website (see <http://www>).

Table 1: Frequency behavior of 13 verbs denoting color in the four corpora of the RNC

Verb	Frequency							
	Main corpus		Newspaper corpus		Oral corpus		Poetic corpus	
	Number of occurrences	NPM	Number of occurrences	NPM	Number of occurrences	NPM	Number of occurrences	NPM
белеть (to whiten)	3356	11,84	159	0,70	20	1,65	897	81,79
чернеть (to blacken)	2968	10,47	134	0,59	6	0,50	733	66,84
краснеть (to redden)	5090	17,96	704	3,08	65	5,37	443	40,39
зеленеть (to turn green)	1948	6,87	184	0,81	12	0,99	501	45,68
желтеть (to turn yellow)	1353	4,77	135	0,59	12	0,99	315	28,72
синеть (to turn blue)	1923	6,78	178	0,78	3	0,25	691	63,01
сереть (to turn gray)	568	2,00	61	0,27	1	0,08	123	11,22
розоветь (to turn pink)	756	2,67	29	0,13	2	0,17	219	19,97
коричневеть (to turn brown)	20	0,07	2	0,01	0	0,00	3	0,27
оранжеветь (to turn orange)	3356	0,01	0	0,00	0	0,00	1	0,09
буреть (to turn brown)	2968	0,52	41	0,18	3	0,25	7	0,64
алеть (to turn scarlet)	5090	2,64	53	0,23	11	0,91	337	30,73
лиловеть (to turn purple)	1948	0,35	1	0,00	0	0,00	45	4,10
Average IPM		5,15		0,57		0,86		30,27

ruscorpora.ru/new/corpora-sem.html).

The purpose of the RNC semantic annotation is to create the ability to search for groups of lexical units that have a common semantic attribute. The RNC semantic annotation is based on the system of classification Russian lexicon, which is adopted in the Lexicograph database [Krasil'shchik, Rakhilina, 1992]. The classification is based on the facet-hierarchical principle, according to which the classification index can combine several features or classes (facets), and each class includes several subclasses of features. To each lexical unit of the corpus a certain class of attributes is assigned in form of an annotation code. Thus the adjective *красный* (red) can be attributed to the class “physical properties”, which includes the sign “color”, the code (tag) `t:physq:color` [Kustova et al., 2005]. Semantic signs are attributed to all word forms of significant words of the corpus.

By an attributed sign or by a class of signs (classification index), a query can be specified and a search can be made to retrieve contexts of words corresponding to a given query.

Each word in the corpus has a full description containing grammatical, semantic and some other features, presented in the form of codes (tags). Descriptions have areas: lemma, grammar, basic semantics, additional semantics and additional attributes. The description is called up by clicking the mouse when the cursor is positioned on the desired word (Fig. 2).

18. Н. Байтов. «Сканер сканирует скатерть...» (2003) [омонимия не снята] [Все примеры](#)

Есть у меня на примете
 добрый приятель один.
 Кликну его в Интернете —
 встретимся и посидим.
 Пусть даже он в бескозырке —
 это всего лишь игра.
 Девушку в **красной** косынке
 нам не забыть нико

красной		
Лемма	красный	
Грамматика	A, f, loc, plen, sg	
Семантика основная	t:physq:color r:qual	

[Н. Байтов. «Сканер сканирует скатерть...» (2003)]

19. Б. Ш. Кенжеев. «Вьется...» (2001-2003) [омонимия не снята]

Но виийствовать —
 так, почти ничего в
 мёртвый мальчик, г
 в чёрно-сахарном п
 то ли в радости ска
 перелевшая лира в
 ты ведь веруешь в т
 ты гуляешь развали
 ты роняешь монетки
 Слезы с потом, как
 льешь?) Какие сух
 звонок череп олеи
 на базальтовой снят
 Раб мой Божий — т
 рабль и непролитое
 серой папиной в пл

Лемма	красный	
Грамматика	A, dat, f, plen, sg	
Семантика основная	t:physq:color r:qual	

Лемма	красный	
Грамматика	A, f, gen, plen, sg	
Семантика основная	t:physq:color r:qual	

Лемма	красный	
Грамматика	A, f, ins, plen, sg	
Семантика основная	t:physq:color r:qual	

Доп. признаки	словарн	
---------------	---------	--

[Б. Ш. Кенжеев. «Вьется...» (2001-2003)]

Figure 2: Description of word form *красной* (red) in the poetry corpus of RNC

When searching, semantic attributes are set in the search query window using the menu tabs that are opened by the link “Semantic signs: choose”. Each menu corresponds to a specific part of speech (see the menu for adjectives, Fig. 3).

The eight menus are provided with a list of semantic attributes for the five parts of speech. Three menus correspond to nouns, for subject, non-subject, and proper names. The semantic

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

Прилагательные

<p>Разряд</p> <input type="checkbox"/> качественные <input type="checkbox"/> относительные <input type="checkbox"/> приклательные <input type="checkbox"/> неизменяемые <p>Таксономия</p> <input type="checkbox"/> размер в том числе: <input type="checkbox"/> большой <input type="checkbox"/> малый <input type="checkbox"/> абсолютный <input type="checkbox"/> расстояние в том числе: <input type="checkbox"/> большое <input type="checkbox"/> малое <input type="checkbox"/> количество в том числе: <input type="checkbox"/> большое <input type="checkbox"/> малое <input type="checkbox"/> абсолютное <input type="checkbox"/> место <input type="checkbox"/> направление	<input type="checkbox"/> время в том числе: <input type="checkbox"/> длительность в том числе: <input type="checkbox"/> большая <input type="checkbox"/> малая <input type="checkbox"/> абсолютная <input type="checkbox"/> возраст в том числе: <input type="checkbox"/> большой <input type="checkbox"/> малый <input type="checkbox"/> абсолютный <input type="checkbox"/> скорость в том числе: <input type="checkbox"/> большая <input type="checkbox"/> малая <input type="checkbox"/> физические свойства в том числе: <input type="checkbox"/> форма <input checked="" type="checkbox"/> цвет <input type="checkbox"/> вкус <input type="checkbox"/> запах <input type="checkbox"/> температура <input type="checkbox"/> качества человека	<input type="checkbox"/> Оценка в том числе: <input type="checkbox"/> положительная <input type="checkbox"/> отрицательная <p>Словообразование</p> <input type="checkbox"/> диминутивы <input type="checkbox"/> аугментативы <input type="checkbox"/> аттенуативы <input type="checkbox"/> забитивы <input type="checkbox"/> каритивы <input type="checkbox"/> потенциальные в том числе: <input type="checkbox"/> возможности <input type="checkbox"/> impossibilities <input type="checkbox"/> отыменные прилагательные в том числе образованные от имен со значениями: <input type="checkbox"/> лица <input type="checkbox"/> животные <input type="checkbox"/> растения <input type="checkbox"/> вещества и материалы <input type="checkbox"/> пространство и место <input type="checkbox"/> здания и сооружения <input type="checkbox"/> инструменты и приспособления <input type="checkbox"/> еда и напитки <input type="checkbox"/> тексты <input type="checkbox"/> части тела и органы <input type="checkbox"/> отглагольные прилагательные <input type="checkbox"/> отадаverbialные прилагательные
---	---	---

Figure 3: The menu for the selection of semantic attributes for adjectives

attribute “color” can be defined for three parts of speech – adjectives, non-subjective nouns and adverbs.

4 Classification of Adjectives with the “Color” Attribute

The texts from the poetry corpus of the RNC for the period from 1750 to 1999. contain 99154 adjectives with the tag “color” (`t:physq:color`). Our verification, however, showed that not every adjective that has or may have a meaning of “color” actually has this tag. It depends largely on which class an adjective belongs to.

In our work [Masevich, Zakharov, 2019], we proposed a classification of color adjectives, which we give below. The classification of colorants is essential for this work, since adjectives of different classes show different quantitative distributions by chronological periods and by authors. Classification is an integral part of the color picture of world.

4.1 Adjectives Meaning Color Per Se

These are adjectives that indicate the main tone (hue), i.e. “the attribute of colors that permits them to be classed as red, yellow, green, blue, or an intermediate between any contiguous pair of these colors” (Webster dictionary). We have revealed 28 of them in the poetry corpus of RNC. Almost all adjectives of this type in the RNC have the tag “color”. The only exception that we were able to identify was the adjective *рдяный* (vermillion, scarlet).

This class includes, first of all, primary colors, or color categories. Different authors

refer to different sets of color designations. We rank to the main adjectives (or basic color categories) a set of twelve lexemes: *белый* (white), *черный* (black), *красный* (red), *желтый* (yellow), *зеленый* (green), *голубой* (light blue), *синий* (blue), *коричневый* (brown), *серый* (gray), *розовый* (pink), *оранжевый* (orange), and *фиолетовый* (purple). Beside adjectives belonging to basic categories, this class includes other adjectives meaning colors per se, e.g. *алый* (scarlet), *лазурный* (azure) etc.

Note. There is a difficulty in describing and classifying Russian terms of color in English (and surely in other languages too), since the system of color nomination and semantics of colors in Russian are very different from English. It causes challenges interpreters have to deal with.

From adjectives of colors per se, as a rule, derivatives are easily formed: a comparative degree, *белее* (whiter or more white), *чернее* (black), *краснее* (approximately, more red); nouns, *белизна* (whiteness), *желтизна* (yellowness), *синева* (approximately, blueness); verbs, *белеть* (whiten), *чернеть* (blacken), *зеленеть* (to turn green), adverbs *черно* (black), *сине* (blue), *бело* (white).

The verbs and nouns of color in English are very few. For example the Russian verb *синеть* according to RNC parallel corpus is translated most frequently as *to turn blue* or less frequently *to be blue*.

However, not all adjectives indicating the main tone possess this property. For example, adjectives such as *лазурный* (azure), *бежевый* (beige), *палевый* (fawn), *фиолетовый* (purple), etc. do not have derivatives (and in some cases, they may have been lost).

When working with terms of color one should remember that the contexts differ when they are used in a direct meaning or in a figurative or phraseological meaning. At this stage of the study, with some exceptions, we do not consider this difference. Quantitative accounting of different contexts is a methodically and technically difficult task that can be solved only by viewing significant arrays of contexts.

In [Masevich, Zakharov, 2019] we showed a significant number of standard collocation and phraseological units with adjectives of color, for example, *белый билет*, *белая гвардия*, *белая баня*, *шито белыми нитками*, *белое духовенство*, *белая ворона*, *белый свет*, *красное лето*, *красная девица*, *красное словцо*, *красное дерево*, *красная строка*, *зеленая скука*, *зеленая улица*, *синий чулок*, *желтый дом*, *желтая пресса*, *желтая лихорадка*, *серые будни*, *серая личность*, *коричневая чума*, *оранжевая революция*, *чёрный день*, *чёрный ход*, *чёрная работа*, *чёрный рынок*, *чёрная зависть*, *розовые надежды*. We also note that such expressions are formed exclusively with adjectives meaning the main color categories.

Note. Some of the expressions one cannot literally translate into English, e.g. the standard collocation *красная девица* literally means *red girl* or *girl of red color* but the real meaning is *beautiful girl*. The others have in English different meaning, e.g. *красная нить* (red string) in Russian means the main topic of something and in English it has several different meanings, one of them is “a thin red string worn to ward off misfortune”.

However, some expressions can be translated literally, e.g. *желтая пресса* (yellow press), *коричневая чума* (brown plague), *оранжевая революция* (orange revolution), *чёрный рынок* (black market).

4.2 “Quasi-Colors” Adjectives

For the convenience of presentation, we have designated a certain type of adjectives, which have tag “color” by the term “quasi-colors”. These are words that characterize a color, but do not mean it, in other words, do not indicate a hue, for example: *темный* (dark), *светлый* (light), *бледный* (pale), *пестрый* (motley), *сумрачный* (somber), *цветной* (colored), *разноцветный* (multi-colored), *бесцветный* (colorless), *пятнистый* (spotty), *клетчатый* (checkered), *блеклый* (faded), *трехцветный* (tricolor), *крапчатый* (speckled) etc. These adjectives describe various visual characteristics of the object – except, perhaps, its shape.

R.M. Frumkina cites the work of Heller and Makris *Parametric Linguistics* [Heller, Makris, 1967] where it is argued that the essential component of color is the basic tone (Eng. *hue*). “Any CN (*color name – auth.*) contains an indication of color in the proper sense (i.e. tone, hue), but there are no CNs that would contain information about brightness and saturation without indicating a color (tone). Indeed, there are no such CNs simply because we do not consider CNs light, bright, dull, brilliant, and motley” [Frumkina, 1984, p. 20].

From a certain point of view, we can talk about a systemic error in the semantic annotation of RNC – assigning the tag “color” to words that do not mean it. However, in the course of the study, the presence of a tag that allows identifying and grouping adjectives with this meaning turned out to be, in our opinion, useful and important for describing the color linguistic picture of the world. In this case, one cannot speak about color names (CNs according to R.M. Frumkina), but about color words, i.e. about words related to color.

Unfortunately, we have to admit to some extent the inconsistency of the semantic mark-up of the RNC. We used synonyms dictionaries, as well as a system for generating semantic fields of color-words based on texts, which allowed us to identify a significant number of words of this type that do not have a semantic marking tag in RNC, for example, *матовый* (matte), *яркий* (bright), *радужный* (rainbow), *узорный* (patterned), *узорчатый* (patterned), *хмурый* (gloomy) etc.

4.3 Analog Adjectives of Color

Some authors call these adjectives the colors of secondary nomination [Dyupina et al., 2013] (relative adjectives). In the studied array of texts from the poetry corpus, only 12 adjectives that have this tag were identified, including: *агатный* (agate), *аметистовый* (amethyst), *белоснежный* (snow-white), *бирюзовый* (turquoise), *дымчатый* (smoky), *золотой* (golden), *изумрудный* (emerald), *малахитовый* (malachite) etc.

Unlike adjectives of colors per se and quasi-colors, analog adjectives, as a rule, do not have a comparative degree and derivatives verbs, adverbs and nouns. However, they can form one of the roots of a complex adjective *алмазно-синий* (diamond blue), *молочно-белый* (milky white), *янтарно-лиловый* (amber purple) etc. They also enter, usually in a metaphoric sense, into stable phrases: *золотые слова* (golden words), *золотое время* (golden time) etc.

At the same time, in addition to the above 12 adjectives, in the RNC poetry corpus we identified about 20 adjectives (more than 1000 usages) of this type that do not have the color tag, e.g. *васильковый* (cornflower), *вишневый* (cherry) etc.

The conclusion, it would seem, is clear: the tag “color” should be ascribed to all analog adjectives. Analysis of the word *малиновый* (raspberry-red) (114 random contexts out of 258

found) showed that in 82 contexts this adjective denotes color.

The remaining uses bear the meanings of “food” (5 contexts), “sound” (9 contexts), “plant” (5 contexts), and metaphorical meaning (13 contexts). However, not every adjective of this class has a dominant color meaning. For such adjectives as *кровавый* (bloody-red), *стальной* (steel-colored), this meaning would be rather peripheral, as a result of which the use of the tag could cause significant information noise. Therefore, the issue, in our opinion, requires further study.

4.4 Adjectives Meaning Specific Colors of Certain Objects (Object-Related)

Object-related adjectives or adjectives with limited cooccurrence range describe animal coloration, the color of human eyes or hair (*карий* – brown, *рыжий* – red) or surface (*крапчатый* – spotty). They are strictly linked to the object, such as a certain animal. e.g. with horse – *гнедой* (bay), *вороной* (raven-black), *буланый* (cream-colored), *саврасый* (dun) and are usually not used with other objects. But there are some counterexamples. Thus, the adjective *незуй* (spotted) one can sometimes use in relation to dogs (like in Chingiz Aitmatov’s novel *Spotted Dog Running along the Bank of the Sea*), as well as to rabbits, deers, birds and even to fishes. We cannot say that the adjective *незуй* (spotted) is used only with living creatures. In poetry, we can see such images as “Дорога пегая – быть теплу” (Н. А. Клюев. “Зима изгрызла бок у стога...”, 1915) ‘the road is spotted, time to be warm’ (Nikolai Klyuchev’s *The Winter Bit the Haystack’s Side*, 1915) or “Талстук вывязан и без спеха – в меру ряб или в меру пег” (Б. А. Слуцкий. “Новый тип”, 1970–1975) ‘the tie is knitted with no hurry, a bit spotted or a bit freckled’ (Boris Slutzky’s *New Type Personality*, 1970–1975). But such word uses are rare and can be explained as metaphors.

The field of use for such adjectives as *крапчатый* (speckled) or *рыжий* (red, red-headed) is much broader. The most frequent bigrams with the adjective *рыжий* are *рыжий мужик* (red-headed man), *рыжий конь* (red horse), *рыжий кот* (red cat) and *рыжий детина* (red-headed guy). These adjectives are also used for descriptions of stock objects (*speckled surface*, *speckled ornament*, *red trousers*, *red polish*). Taking this into account with the fact that the adjective *рыжий* (red, red-haired) shows the hue, we classified it as a color per se, not as an object-related adjective, while the adjective *крапчатый* (speckled), which does not show the hue, we classified as a quasi-color.

4.5 Derivatives of Color Adjectives

4.5.1 Diminutives

This class includes such adjectives as *беленький* (diminutive of white), *желтенький* (diminutive of yellow) etc. and quasi-color derivatives such as *пестренький* (diminutive of motley) etc. The adjectives in this form are practically untranslatable into English, for they do not mean the color of thing only, but to a bigger degree the psychological attitude to it, from tenderness and sympathy to condescension and even light scorn.

4.5.2 Adjectives of Likeness

Here, we mean adjectives like *зеленоватый* (similar to green, greenish), *желтоватый* (similar to yellow, yellowish), *красноватый* (similar to red, reddish), *темноватый* (rather dark), *бледноватый* (a bit pale). Like diminutives adjectives, these adjectives do not have an exact English equivalent and mean likeness to a certain color (or resemble a certain color). We include in this class also verbal adjectives, which show the result of some process. These are words like *посинелый* (became blue), *почернелый* (turned black), *потускнелый* (became lackluster).

4.5.3 Double (Compound) Color Adjectives

As we have already noticed in the previous study [Masevich, Zakharov, 2019], double (as well as triple and more) adjectives are the most numerous group, but every word has low frequency. It should be noticed that the RNC tags these adjectives according to two random rules: in some cases, a compound adjective is observed as one lexical item, i.e. *чёрно-белый* (black-and-white), *тёмно-синий* (dark blue), *нежно-голубой* (light blue) tagged as one word of color. Searching by the tag “color” in the RNC, we could find 56 of such units. In other cases, such compound adjectives are tagged as two separate lexical units, e.g. *ярко-красный* (bright red) is tagged as two lexical items *яркий* (bright) and *красный* (red). Using the random review, we could make a list (and probably not a full one) of 740 adjectives of color tagged as two words with separate descriptions.

5 Research Technique

Step 1. Within the poetic corpus of the RNC we have created 25 sub-corpora, each of which covered 10 years in accordance to the time of writing the texts. The whole set covered the period from 1750 to 1999.

In every sub-corpus we made a search of adjectives by the semantic tag “color” (Fig. 3).

In each set of search results, we fixed the following data:

the volume of the user sub-corpus (the total number of words);

the number of texts (poems) which include tagged adjectives;

the number of tagged adjectives;

the list of lemmas on every page of the search results (Table 2).

Lists of lemmas from all search result pages were copied to a united MS Excel table sheet (Fig. 4).

By functions of MS Excel, the summary table of value was formed, in which the sum of uses (i.e. the sum of page-to-page results is made) of each lemma was counted (Fig. 5).

To the table got from the summary table the IPM column was added. The rate of IPM is counted using the formula $b/c \times 1000000$ where b is the number of uses of a lexeme and c is the volume (the quantity of words) of the sub-corpus (for given in the corpus of texts from 1900 to 1909). The formula was inserted in the string of formulas, with the results generated into

Lemma	Frequency
1 лемма	c/y
2 светлый	28
3 черный	22
4 белый	18
5 голубой	13
6 красный	12
7 темный	11
8 зеленый	10
9 синий	9
10 розовый	9
11 серый	8
12 бледный	7
13 цветной	6
14 бирюзовый	6
15 пестрый	5
16 лазурный	5
17 золотистый	5
18 мутный	4
19 багряный	4
20 багровый	4
21 алый	4
22 оранжевый	3
23 сумрачный	2
24 сизый	2
25 изумрудный	2
26 черноголовый	1

Figure 4: A fragment of the overall list of page-to-page lemma lists of search results for the tag “color” within the 1900–1909 sub-corpus

Lemma	Sum of use frequency
4 агатовый	6
5 аленький	11
6 алый	466
7 аметистовый	2
8 багровый	72
9 багряный	101
10 белеший	1
11 беленский	7
12 белесоватый	3
13 белесый	6
14 белоснежный	41
15 белый	1620
16 бесцветный	18
17 бирюзовый	33
18 бледно-голубой	2
19 бледный	824
20 буланный	1
21 бурый	27
22 вороной	6
23 гнедой	4

Figure 5: A fragment of the summary table of value with the sum of use frequency for each lemma for the period 1900–1909

the table cells automatically. The procedure is finished with inserting the “period” column

Table 2: A fragment of the list of lemmas on the page with search results from the RNC

Lemma	Number of word uses
белый / white	19
черный / black	14
желтый / yellow	12
темный / dark	10
светлый / light	9
красный / red	6
...	...

C2 =B2/858144*1000000

1900-09ПозтичКорпусЛеммы.xlsx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Лемма	с/у	IPM	Период										
2	белый	1620	1887,80	1900-е гг.						Период 1900 -1909				
3	темный	1277	1488,10	1900-е гг.										
4	черный	1038	1209,59	1900-е гг.										
5	светлый	993	1157,15	1900-е гг.										
6	бледный	824	960,21	1900-е гг.										
7	красный	748	871,65	1900-е гг.										
8	синий	682	794,74	1900-е гг.										
9	зеленый	655	763,28	1900-е гг.										
10	голубой	590	687,53	1900-е гг.										
11	алый	466	543,03	1900-е гг.										
12	серый	360	419,51	1900-е гг.										
13	желтый	317	369,40	1900-е гг.										
14	розовый	261	304,14	1900-е гг.										
15	лазурный	197	229,57	1900-е гг.										
16	сумрачный	152	177,13	1900-е гг.										
17	золотистый	136	158,48	1900-е гг.										
18	пестрый	121	141,00	1900-е гг.										
19	лиловый	111	129,35	1900-е гг.										
20	сизый	111	129,35	1900-е гг.										
21	серебристый	109	127,02	1900-е гг.										
22	багряный	101	117,70	1900-е гг.										
23	золотой	99	115,37	1900-е гг.										

Figure 6: A fragment of the frequency table for color adjectives in the sub-corpus of texts written from 1900 to 1909

containing the signification of the observed lapse of time (Fig. 6).

Describing the first step, we should notice that there is a certain difficulty when defining the sub-corpus, as the creation dates of many texts are not exactly defined and shown as the time lapse. If the time lapse of a text creation is defined, say, as 1909–1911, it can be included both in the massive of 1900’s and in the massive of the 1910’s. We cannot be sure about the quantity of lexical units within our sets, which are reflected twice, but we hope that it had no significant influence on results.

Step 2. 25 tables created at the first stage (according to the number of decades) are united into an overall MS Excel table. In our case, the overall table (Fig. 7) contains 2,004 rows (excluding the headline). The data from the table may be sorted by one or several criteria. We can also build the pivot diagrams basing on the table, thus, some regularities of frequency behavior of certain lexical units can be observed.

Then, to every row of the overall table we added the letter code of the class from the classification described above (Fig. 8). The meanings of the codes are as follows: AH – analog adjectives, ДВ – double adjectives, ДМ – diminutives, КВ – quasi-colors, ПЛ – adjectives of likeness, Пр – object-related adjectives, ЦЦ – adjectives of color per se. We simplified a bit the routine work of manual inserting the codes, but we do not describe here the details.

Step 3. From the table (Fig. 8), we generated a pivot diagram, with period as filter, classes and lemmas (order of terms is of importance!) as categories and IPM as values (Fig. 9).

MS Excel allows the setting of the different parameters for getting the final data. For example, we can set and get the diagram of use of adjectives of all classes during the whole studied time lapse (1750–1999).

It is obvious that some data are incomparably smaller than others, so the corresponding elements of the diagram would not be seen. In such cases, two methods can be used, depending

on the problems. Firstly, we can make a logarithmic scale on the vertical axis (Fig. 10). Secondly, we can create the diagrams selecting the comparable data.

6 Results of the Research

6.1 Frequency Behavior of The Color Adjectives with the Time Lapse

We can see from the diagram (Fig. 10) that, despite the uncomfortable scale due to the big data, the quantitative difference in coloratives of different classes is obvious. Most of the lexical units are double adjectives.

According to the diagram, the number of lexical units for analog adjectives is less than for all others, but it is obvious that it is because only about one-third of the coloratives belonging to such class have signs of semantic annotation.

At the same time, such value as IPM doesn't depend on the consistency of assignment of tags or the bigger or smaller number of lexical units. Therefore, the diagram shows the objective value of the frequency behavior of words. The adjectives meaning the color per se have the highest IPM. The IPM of quasi-color and analog adjectives is also quite high (Fig. 11). From this diagram (Fig. 11), we can see the proportions between the frequency of adjectives of color of three most frequent classes. The adjectives of other classes have much lower frequency, incomparable to these three classes.

Comparing the IPM change dynamics for coloratives from different classes within the time lapse (Fig. 12), we can see very interesting tendencies. In the 18th century, the analog adjectives were the most frequent. In the 19th century, the quasi-colors became the most

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	лемма	с/у	ipm	период									
2	златой	49	400,25	1750-е гг.									
3	красный	39	318,56	1750-е гг.									
4	светлый	19	155,20	1750-е гг.									
5	зеленый	16	130,69	1750-е гг.									
6	белый	8	65,35	1750-е гг.									
7	темный	7	57,18	1750-е гг.									
8	черный	5	40,84	1750-е гг.									
9	бледный	5	40,84	1750-е гг.									
10	багряный	5	40,84	1750-е гг.									
11	желтый	3	24,50	1750-е гг.									
12	багровый	3	24,50	1750-е гг.									
13	серый	2	16,34	1750-е гг.									
14	румяный	2	16,34	1750-е гг.									
15	рыжий	1	8,17	1750-е гг.									
16	розовый	1	8,17	1750-е гг.									
17	разноцветный	1	8,17	1750-е гг.									
18	пегий	1	8,17	1750-е гг.									
19	мутный	1	8,17	1750-е гг.									
20	златой	70	395,52	1760-е гг.									
21	красный	26	146,91	1760-е гг.									
22	черный	26	146,91	1760-е гг.									
23	светлый	23	129,96	1760-е гг.									

Figure 7: A fragment of the overall table, sorted by period (ascending) and IPM (descending)

	A	B	C	D	E	F	G	H	I	J
1	Лемма	С/у	IPM	Период	Класс					
2	агатовый	1	1,85	1830-е гг.	АН					
3	агатовый	1	2,46	1850-е гг.	АН					
4	агатовый	1	2,50	1860-е гг.	АН					
5	агатовый	6	6,99	1900-е гг.	АН					
6	агатовый	1	0,72	1910-е гг.	АН					
7	агатовый	4	2,43	1920-е гг.	АН					
8	агатовый	1	0,87	1930-е гг.	АН					
9	агатовый	1	1,64	1950-е гг.	АН					
10	агатовый	1	1,85	1960-е гг.	АН					
11	агатовый	1	14,67	1990-е гг.	АН					
12	аленький	1	3,93	1790-е гг.	Дм					
13	аленький	1	2,26	1840-е гг.	Дм					
14	аленький	1	2,50	1860-е гг.	Дм					
15	аленький	11	12,82	1900-е гг.	Дм					
16	аленький	10	7,19	1910-е гг.	Дм					
17	аленький	3	1,82	1920-е гг.	Дм					
18	аленький	5	4,36	1930-е гг.	Дм					
19	аленький	1	1,34	1940-е гг.	Дм					
20	аленький	1	1,64	1950-е гг.	Дм					
21	аленький	3	5,54	1960-е гг.	Дм					
22	алый	2	10,78	1770-е гг.	СЦ					
23	алый	6	37,87	1780-е гг.	СЦ					
24	алый	37	125,78	1790-е гг.	СЦ					

Figure 8: Fragment of the overall table with the added column “class” (sorted by lemmas)

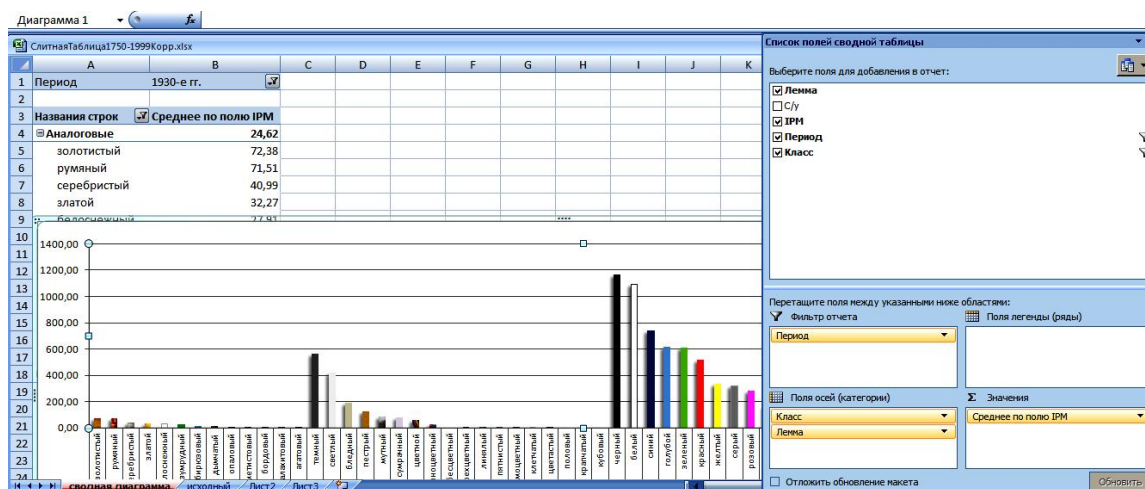


Figure 9: Building of the pivot diagram for changing of the IPM dynamics by classes during the certain time lapse (function of MS Excel)

frequent ones. In the 20th century, the adjectives of the color per se had the highest IPM. We can also notice that the highest rates of the adjectives of quasi-colors and colors per se are

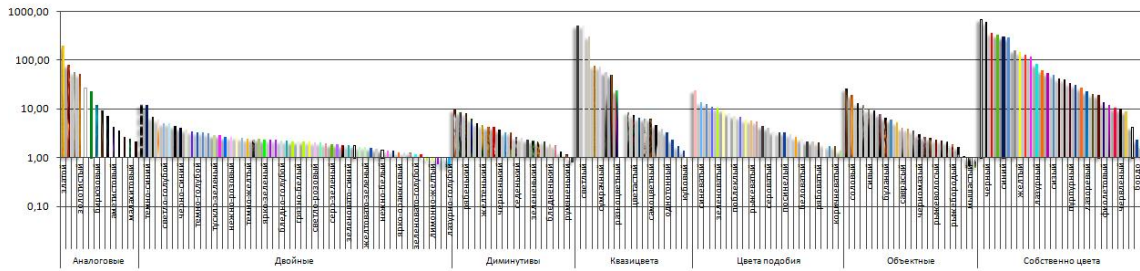


Figure 10: The pivot diagram of use of the adjectives of for whole period (1750–1999) and all classes with the logarithmic scale on the vertical axis (IPM)

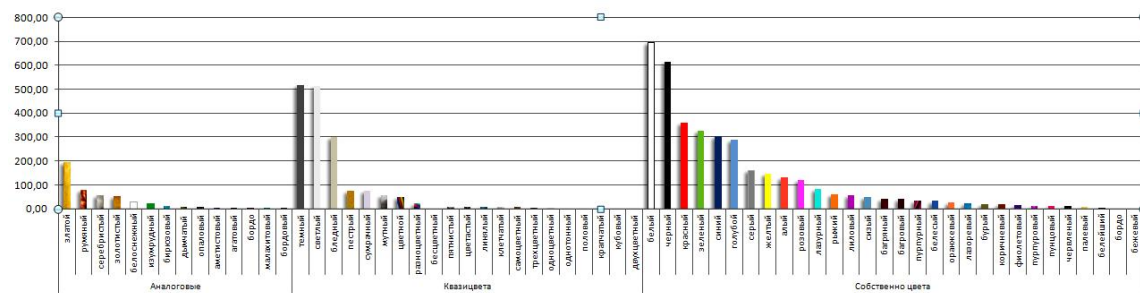


Figure 11: A diagram for analogs, quasi-colors and colors per se with the absolute scale on the vertical axis (IPM)

identified during the last decade of the 19th century (quasi-colors) and the first decade of the 20th century (colors per se). After that, the IPM for both classes was falling, much more for quasi-color adjectives until the middle of the century.

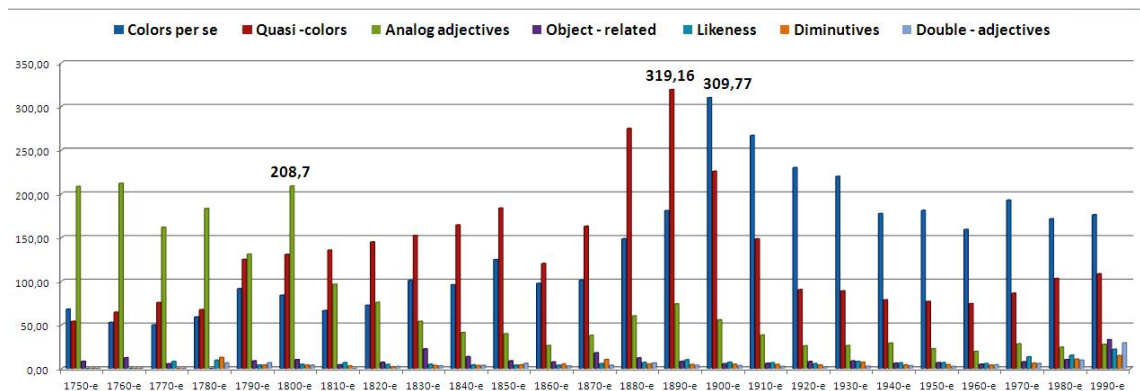


Figure 12: The dynamics of the IPM change for classes of color adjectives within 1750 to 1999 period

We'll now construct diagrams that allow us to see the frequency behavior of lexemes of adjectives over time periods of 18th, 19th and 20th centuries, inside three classes of most frequent color adjectives: analog adjectives, quasi-colors, and adjectives per se.

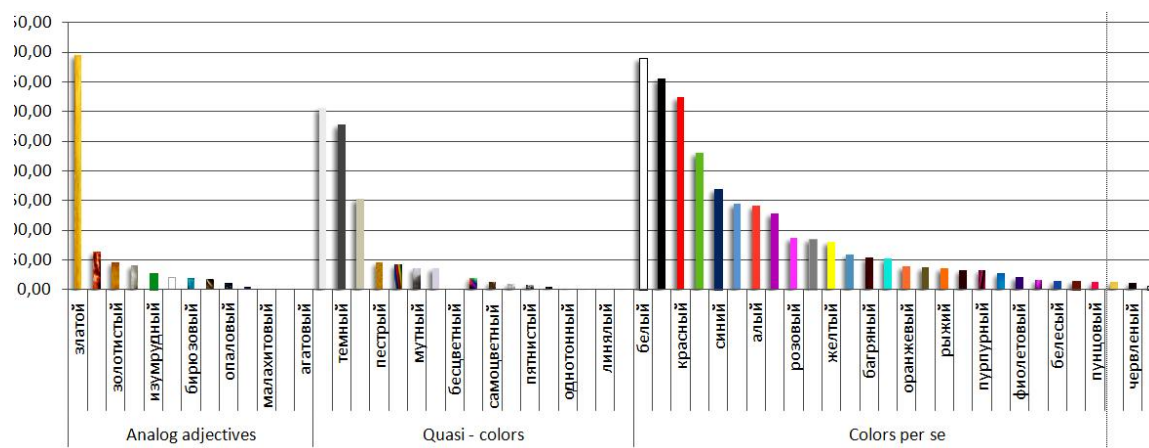


Figure 13: IPM of adjectives with the tag “color”, sorted by class and then by descending IPM for 1750–1799

The diagram shows that the high frequency of analog adjectives is mainly due to the high frequency of the adjective *златой*, which is the archaic form of the word *золотой* (golden). The IPM of this lexical unit is more than twice the IPM of the most frequent adjectives belonging to the different classes: adjective colors per se and quasi-colors. Apparently, the point is the polysemy of this word. Besides the meaning of color, it, as is characteristic of an analog adjective, means made of gold, but also has a very common metaphorical meaning ‘happy’. It is important to note that the maximal IPM in this chronological period is close to 400 – the adjective *златой* (Fig. 13).

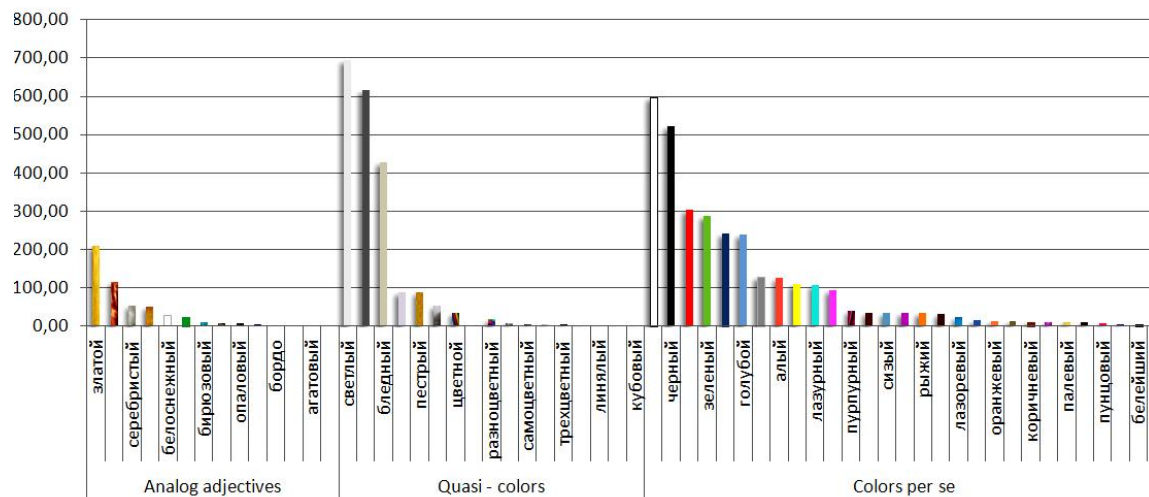


Figure 14: IPM of adjectives with the tag “color”, sorted by class and then by descending IPM for 1800–1899 period

In the 19th century, as can be seen in Fig. 14, the rise of IPM is observed in quasi-colors, with the three most frequent lexemes of this class being adjectives *светлый* (light), *темный* (dark) and *бледный* (pale). Within the class of colors per se, the most frequent lexemes are *черный* (black) and *белый* (white), which corresponds to the conclusions of our previous

publications [Masevich, Zakharov, 2019]. The highest IPM in this period is close to 700 – the adjective *светлый* (light).

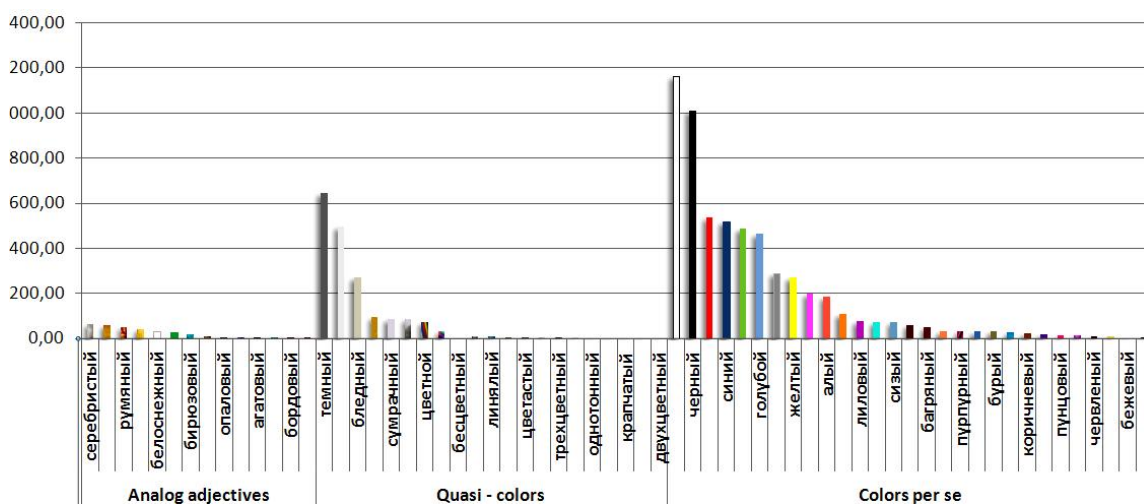


Figure 15: IPM of adjectives with the tag “color”, sorted by class and then by descending IPM for 1900–1999 period

As one can see in Fig.15, the most frequent in the poetic texts of the 20th century are the adjectives of color per se. Moreover, the highest values of IPM are observed in the lexemes *белый* (white) and *черный* (black), i.e. achromatic colors. Note that the IPM value of the adjective *темный* (dark) is higher than that of the adjectives *красный* (red), *синий* (blue), *зеленый* (green), *голубой* (light blue), and the IPM of *светлый* (light) is approximately equal to the average of these four words. The highest IPM in the texts of 20th century is close to 1200.

7 Conclusion

The adjectives that we examined in this article have one common attribute – the presence of the color tag in the semantic annotation of RNC. According to the data obtained, the frequency of occurrence of adjectives in a poetry corpus is higher than their frequency in the main corpus and is noticeably higher than in oral speech or newspaper articles.

The description of the color picture of the world in Russian poetic texts is presented in terms of lexemes and semantic classes. We have divided adjectives of colors into five groups, which give some structure to the color world picture.

A characteristic feature of the frequency behavior of color adjectives in general is a pronounced growth at the end of the 19th century and the beginning of the 20th.

In our opinion, we have been able to identify the main models of adjectives frequency behavior in a rather significant time period. It is shown that in the second half of the 18th century, the analog adjectives turn out to be the most frequent. It should also be noted that in 19th century poetic texts the frequency of adjectives of quasi-colors is higher than of adjectives of color per se. In 20th century texts adjectives of color per se become most frequent.

It is noteworthy that, along with a relatively small number of the most frequent adjectives,

there is, a very large number (nearly 800) of lexemes of complex color terms of a single use. This structure of the lexical array resembles the frequency distribution according to Zipf's law.

The study reveals significant shortcomings of RNC semantic annotation, its inconsistency, which does not allow to take into account the real number of lexemes of some adjective types, in particular those that we name analog, quasi-color, as well as double adjectives.

At the next stage of our work we plan to supplement our studies with a detailed analysis of the behavior of low-frequency adjectives classes, as well as some data on the use of color terms in the vocabularies of individual authors.

As a rule, we refrain from interpreting the data obtained. Nevertheless, it seems that our work results, in addition to the statistical data presented, could be proposed as a methodological approach to research based on a linguistic corpus.

References

- [Vasilevich et al., 2005] Vasilevich A.P., Kuznetsova S.N., Mishchenko S.S. *Color and color names in Russian* / Under the general ed. of Vasilevich A.P. M.: Kom Kniga, 2005. P. 216 (In Rus.) = Vasilevich A.P., Kuznecova S.N., Mishchenko S.S. *Cvet i nazvaniya cveta v russkom yazyke* / Pod obshch. red. A. P. Vasilevicha. M.: Kom Kniga, 2005. S. 216.
- [Gerasimov, 1969] Gerasimov G.I. *Color adjectives* // Russkij yazyk za rubezhom. 1969. № 3. P. 26–35. (In Rus.) = Gerasimov G.I. *Prilagatel'nye, oboznachayushchie cvet* // Russkij yazyk za rubezhom. 1969. № 3. S. 26–35.
- [Grishina et al., 2006] Grishina E.A., Korchagin K.M., Plungyan V.A., Sichinava D.V. *Poetic corpus within the framework of the NCR: general structure and prospects for use* // Nacional'nyj korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy. SPb.: Nestor-Istoriya. 2009. P. 71–113. (In Rus.) = Grishina E.A., Korchagin K.M., Plungyan V.A., Sichinava D.V. *Poeticheskij korpus v ramkah NKRYa: obshchaya struktura i perspektivy ispol'zovaniya* // Nacional'nyj korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy. SPb.: Nestor-Istoriya. 2009. S. 71–113.
- [Dyupina et al., 2013] Dyupina Yu.V., Shakirova T.V., Chumanova N.A. *Color classification in linguistic literature* // Molodoj uchenyj. 2013. № 1. P. 220–221. URL: <https://moluch.ru/archive/48/6090/> (last access: 09.07.2019). (In Rus.) = Dyupina Yu.V., Shakirova T.V., Chumanova N.A. *Klassifikacii cvetooboznachenij v lingvisticheskoj literature* // Molodoj uchenyj. 2013. № 1. P. 220–221. URL: <https://moluch.ru/archive/48/6090/> (last access: 09.07.2019).
- [Masevich, Zakharov, 2019] Zakharov V.P., Masevich A.Ts. *Frequency behavior of color adjectives in Russian poetic texts* // Vestnik NGU. Seriya: Lingvistika i mezhkul'turnaya kommunikaciya. 2019. Vol. 17, № 1. P. 21–48. (In Rus.) = Zakharov V.P., Masevich A.C. *Chastotnoe povedenie prilagatel'nyh cveta v russkih poeticheskikh tekstah* // Vestnik NGU. Seriya: Lingvistika i mezhkul'turnaya kommunikaciya. 2019. Tom 17, № 1. S. 21–48.

- [Kezina, Perfilova, 2015] Kezina S.V., Perfilova M.N. *The etymology of color designations of archaisms in the Russian language (according to the dictionary of the Russian language of the 11th–17th centuries)* // Vestnik Penzenskogo gosudarstvennogo universiteta № 4 (12), 2015. P. 19–23. (In Rus.) = Kezina S.V., Perfilova M.N. *Etimologiya cvetooboznachenij arhaizmov v russkom yazyke (po slovarju russkogo yazyka XI–XVII vv.)* // Vestnik Penzenskogo gosudarstvennogo universiteta № 4 (12), 2015. S. 19–23.
- [Krasil'shchik, Rakhilina, 1992] Krasil'shchik I.S., Rakhilina E.V. *Subject Names in Lexicograph Systems* // NTI. Ser. 2. 1992. № 9. P. 24–31. (In Rus.) = Krasil'shchik I.S., Rakhilina E.V. *Predmetnye imena v sisteme "Leksikograf"* // NTI. Ser. 2. 1992. № 9. S. 24–31.
- [Kustova et al., 2005] Kustova G.I., Lyashevskaya O.N., Paducheva E.V., Rakhilina E.V. *Semantic annotation of the vocabulary in the National corpus of the Russian language: principles, problems, prospects* // Nacional'nyj korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektivy. M., 2005. P. 155–174. (In Rus.) = Kustova G.I., Lyashevskaya O.N., Paducheva E.V., Rakhilina E.V. *Semanticheskaya razmetka leksiki v Nacional'nom korpusse russkogo yazyka: principy, problemy, perspektivy* // Nacional'nyj korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektivy. M., 2005. S. 155–174.
- [Frumkina, 1984] Frumkina R.M. *Color, sense, similarity (aspects of psycholinguistic analysis)*. M., 1984. (In Rus.) = Frumkina R.M. *Cvet, smysl, skhodstvo (aspekty psiholingvističeskogo analiza)*. M., 1984.
- [Berlin, Kay, 1969] Berlin B., Kay P. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press, 1969.
- [Heller, Makris, 1967] Heller L., Makris J. *Parametric Linguistics*. The Hague, Paris, 1967.
- [Kay, Regier, 2006] Kay P., Regier T. *Language, Thought and Color: Recent Developments*. Trends in Cognitive Sciences, 10 (2), 2006. P. 51–54.
- [Kay, McDaniel, 1978] Kay P., McDaniel Ch.K. *The Linguistic Significance of Meanings of Basic Color Terms*. Language, 54 (3), 1978. P. 610–646.