

Lexical Diversity and Colour Hues in Russian Poetry: A Corpus-Based Study of Adjectives

Olga Lyashevskaya

National Research University Higher
School of Economics; Vinogradov
Institute of the Russian Language RAS
(Moscow, Russia)
olesar@yandex.ru

Ekaterina Vlasova

National Research University Higher
School of Economics (Moscow, Russia)
evlasova@hse.ru

Kristina Litvintseva

National Research University Higher School of Economics (Moscow, Russia)
tinalitvina@gmail.com

Abstract

This paper describes the distribution of colour adjectives in Russian poetry of the Silver Age and defines individual preferences with regard to poetic tradition, syllable structure, and metrical restrictions. The research method combines a lexico-semantic approach, formal literary analysis, and quantitative metrics obtained via the frequency database of the Russian Poetry Corpus (over 10 M words, incl. 1 M adjectives). The database allows the user to compare subcorpora and create graphs of timeline distribution, which demonstrate that the lexical diversity and relative frequencies of colour adjectives start to grow rapidly in the 1890s, as modernists employ colour adjectives to upgrade the poetic inventory. The adjectives referring to non-banal hues (e.g. fioletovyy ‘violet’, lazorevyy ‘azur’) belong to the middle part of the ranked wordlist. Correspondence analysis of the data reveals individual colour preferences and stylistic similarities among the most prominent poets of the Silver Age; for example, Anna Akhmatova and Alexander Blok are similar regarding their use of the white hues. The distribution of the selected colour hue adjectives across metrical types highlights the strong association of multi-syllabic adjectives with certain meters, although some words have a more complex distribution.

1 Introduction

Slavic quantitative corpus linguistics (Kopotev et al. 2018; Divjak et al. 2017) and formal studies of poetry in general (Dresher–Friedberg 2008) have elaborated a group of computational methods and tools applicable to Russian versification, which has always heavily relied on statistical data as the basis for generalizations regarding meter, rhyme, and other formal and linguistic features of poetic language (see Gasparov 2005; Taranovsky 2010; Jakobson 1973; Yarkho 2006, to name only a few; see also overviews in Kizhner et al. 2018).

As quantitative analysis requires processing a large collection of texts, linguists responded to this challenge by creating the Poetry Corpus as a part of the Russian National Corpus. The Russian Poetry Corpus is a digital open-access resource provided with the standard morphological and lexico-semantic tagging and a number of specific tags particularly suited for poetic language. For example, the search options offer possibilities to collect texts written in various poetic meters, genres, certain patterns of rhyme, verse forms, and even graphical shapes. For more information about the preparation of the text collection included in the Poetry Corpus and the principles of its annotation see (Grishina et al. 2009).

The Russian Poetry Corpus has proven to be an effective source for rapid extraction of the raw and normalized frequencies required for the stylistic and diachronic research of poetic language. As a digital resource it provides additional large-scale data for verification and support of traditional close-reading methods. Comprising more than ten million tokens with multilevel annotation, the Russian Poetry Corpus is already a large representative resource for quantitative studies, including digital literary studies and computational stylistics (Jacobs 2018). However, for revealing more sophisticated patterns within the data, the text collection requires additional annotation of poetic and linguistic features, as well as tagging of relevant historical background information essential for observation of cultural trends.

This article describes a new resource assembled from the data and annotation of the Russian Poetry Corpus, henceforth called a frequency database of the corpus. The new tool has been designed by an interdisciplinary research group from the Higher School of Economics (Moscow). This project aims to design a database with enhanced and elaborated annotation and create an open-access web application with statistical tools for data summarising, filtering, and pattern structuring; for more details see (Lyashevskaya et al. 2018).

This paper contains explanatory case studies of lexical diversity in the Russian poetic tradition with regard to the evolution of poetic vocabulary, as well as word syllable structure and metrical restrictions. Case 1 is a comparative study supported by a correspondence analysis with a focus on individual similarities and distinctions between the most prominent poets of the Silver Age. Case 2 examines correlations between the syllable structure of colour hue adjectives and types of meter. The combination of prosodic and lexical analysis reveals that poets' lexical choices were not random, but needed to fit the accentual-syllabic scheme of a verse line. Therefore, multi-syllabic adjectives, such as *izumrudnyi*, demonstrate a strong association with certain types of trisyllables and disyllables, producing the effect of a 'colour halo'.

2 The colour hue adjectives in Silver Age Poetry

The following example illustrates how the frequency database can be used for the analysis of lexical diversity and a poet's word choices. During the Silver Age a number of poetic groups established a new principle of aesthetics and art syncretism, trying to combine painting and poetry. By exploring colour hue adjectives, this case study aims to reveal how the aesthetic rationale influenced the poetic lexicon. This is done by applying several methods. The first method involves a small-scale diachronic analysis of word frequencies during the 19th-20th centuries. At the next stage, we apply a method of correspondence analysis (CA) to define frequency-based associations between colour hue adjectives and certain poets. The CA method also involves clustering poets based on the links between words and authors.

At the preparatory stage, we extracted a frequency list of the colour adjectives, using the lexico-semantic annotation of the frequency database. Then we compiled a list of adjectives for colour hues by filtering out the most frequent lexemes (such as красный 'red', синий 'blue') and hapax legomena (such as алмазно-рубиновый 'diamond ruby'). The middle part of the frequency list consists of the following lexemes referring to non-banal hues and subject to further analysis: фиолетовый 'violet', лиловый 'lilac', лазурный 'azure', багряный 'blood-red', пурпурный 'tyrian purple', белоснежный 'snow-white', изумрудный 'emerald', лазоревый 'azure', бирюзовый 'turquoise', золотой 'golden-yellow', сумрачный 'murky'.

These lexemes occur more than 100 times each and, apart from the most frequent colour hue adjectives (белый 'white', черный 'black', темный 'dark', светлый 'light', красный 'red'), do not constitute idiomatic collocations (красная армия 'the red army', белое вино 'white wine'). The most frequent lexemes are usually distributed equally over texts regardless of individual and genre variation. The less frequent words have the potential to become a stylistic feature of a personal style as well as of an epoch.

The micro-diachronic research comprises the period from 1801 to 1970, including several decades before and after the Silver Age. The graph shows that the frequency of the adjective фиолетовый 'violet' starts to increase rapidly from the 1880s to the 1920s. During this time, its frequency increases from ~3 ipm to ~21 ipm and then remains at this rate. The first poet who brought the adjective фиолетовый 'violet' into poetry was Vasily Zhukovsky. However, the frequency of this lexeme reached its peak only in the Silver Age. Apart from Zhukovsky, фиолетовый 'violet' occurred in poetry only three times. In 1895, Valery Bryusov used the collocation фиолетовые руки 'violet hands'; later on, the adjective фиолетовый 'violet' occurs in poems of Maksimilian Voloshin, Andrei Bely, Vyacheslav Ivanov, Ivan Bunin, Aleksandr Blok and many others. This small-scale diachronic frequency analysis has demonstrated that фиолетовый 'violet' is a specific stylistic feature of the Silver Age.

Alongside the interest in the violet colour, poets employ adjectives denoting its hues. For example, the adjective лиловый 'lilac' has a similar diachronic graph as фиолетовый 'violet' (FIG. 1).

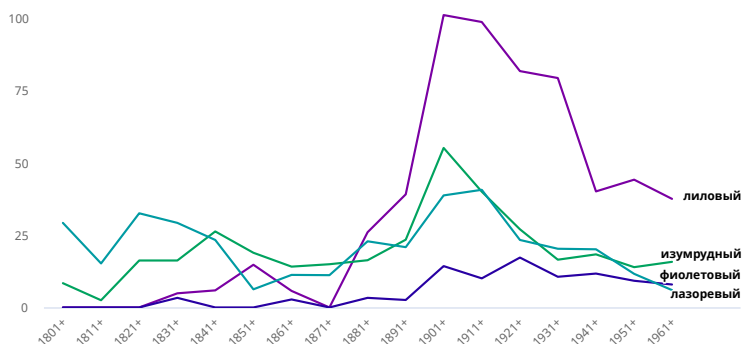


FIG. 1: Occurrences of the adjectives лиловый ‘lilac’, фиолетовый ‘violet’, изумрудный ‘emerald’, лазоревый ‘azure’ by decade, in ipm

As the Russian Poetry Corpus shows, the frequency of fluctuates at the rate from ~1 ipm to ~10 ipm. Over the 50 years from the 1870s to the 1920s, its frequency increases to ~120 ipm, stays at this rate for about 10 years, and then starts gradually declining in the 20th century.

How does this fashion for certain words emerge and do certain poets play a role in this process? Why does the frequency of some colour hue adjectives increase and decrease rapidly? The toolkit of the frequency database allows one not only to extract generalized frequency data across decades, but also to explore frequency distributions within the corpora of certain poets. For example, the search results show that лиловый ‘lilac’ is regularly attested in the poems of Ivan Bunin, Vyacheslav Ivanov, Mirra Lokhvitskaya, Boris Pasternak, and Igor Severyanin. These poets contribute most to the high frequencies of лиловый ‘lilac’ in the Silver Age.

Two other two highly frequent colour hue adjectives of the Silver Age are лазоревый ‘azure’ and изумрудный ‘emerald’. The frequency of лазоревый ‘azure’ during the period of 1890-1930 does not decrease below 34 ipm, and the average frequency of this adjective is two times higher than after the Silver Age. The adjective изумрудный ‘emerald’ has a similar diachronic distribution. Its lowest frequency within this period is about 50 ipm, and this is twice as much as the Silver Age. At the next stage, we visualized distributional data drawn from the database using the CA method (Levshina 2015; Kassambara 2017) as applied to the use of the colour adjectives in focus by individual authors.

For a case study, we took nine subcorpora written by Valery Bryusov, Alexander Blok, Konstantin Bal’mont, Igor Severyanin, Nikolay Gumilev, Anna Akhmatova, Marina Tsvetaeva, Osip Mandel’shtam, and Boris Pasternak. Texts were chosen without regard to their date of composition. For its source data, CA takes a contingency table that shows how the linguistic units (9 adjectives of colour hues, in our case) are distributed across the subcorpora (nine authors, in our case). The distribution of each adjective across the subcorpora we call a colour profile, and the distribution of the uses of the adjectives by each author we call an author profile. First, we calculated an

average profile for both adjectives and authors. Second, we computed the distance between each pair of colour profiles and from each colour profile to the average colour profile. The distances for the author profiles are calculated the same way. In addition, a matrix of distances is plotted onto the 2D space using the method of multi-dimensional reduction. The closer the data points are on the horizontal or the vertical axes, the closer are their profiles. The closer they are to the origin (0,0), the closer their profiles to the average profile.

FIG. 2 illustrates the similarity among the adjectives of colour as measured by their frequency distributions in subcorpora. The profiles of белоснежный 'snow-white', лазурный 'azure', золотой 'golden-yellow', изумрудный 'emerald', and самоцветный 'semiprecious' can be regarded as opposed to the profiles of пурпурный 'tyrian purple', бирюзовый 'turquoise', фиолетовый 'violet', багряный 'blood-red', сумрачный 'murky', лазоревый 'azure', and лиловый 'lilac' (top vs. bottom part of the plot); while the profiles of багряный 'blood-red', сумрачный 'murky' and the profiles of бирюзовый 'turquoise', фиолетовый 'violet' are two poles on the horizontal axis (left vs. the right part of the plot). Furthermore, the profiles of изумрудный 'emerald' and самоцветный 'semiprecious' are much closer to the origin than the profile of белоснежный 'snow-white'. This can be interpreted to mean that изумрудный 'emerald' and самоцветный 'semiprecious' are used roughly equally by the different authors, whereas белоснежный 'snow-white' is used considerably more frequently in one or several subcorpora than in others. The axis labels provide information as to the extent to which the variance in the frequency profiles is explained by the 2D visualization—in other words, how much information was lost when the multidimensional space was reduced to two dimensions ($100\% - 44.9\% - 25.3\% = 29.8\%$).

The authors' profiles can be plotted the same way. FIG. 3 shows a global pattern within the data (symmetric biplot), the colours' profiles (blue points) and the authors' profiles (red triangles) being plotted simultaneously.

The plots on FIG. 2 and 3 were formed using the subcorpora of nine authors (all except Pasternak). The reason is that his profile differs greatly from that of all other poets, so a user would see a dense cloud of points in the center and an outlier. The technique of supplementary points allows one to plot the outlier's point over the plot created for the rest of the data. In other words, the colour profiles of Pasternak and Severyanin are not particularly similar, but they are more similar than the profiles of Pasternak and Bryusov, or Pasternak and Tsvetaeva.

As expected, the graph shows that both Boris Pasternak and Igor Severyanin use the adjective лазоревый 'azure' and лиловый 'lilac' frequently. The visualization illustrates as well that the lexemes бирюзовый 'turquoise' and фиолетовый 'violet' also belong to their poetic lexicon. Despite the low average frequency of colour adjectives in the poetry of Nikolay Gumilev and Valery Bryusov, the graph demonstrates that their poems have similar lexical features, such as the colour terms of багряный 'blood-red', пурпурный 'tyrian purple', and сумрачный 'murky'. Anna Akhmatova stands out from the other poets, in that her poetry is imbued with colour adjectives. She also used the highest frequency words, such as белоснежный 'snow-white',

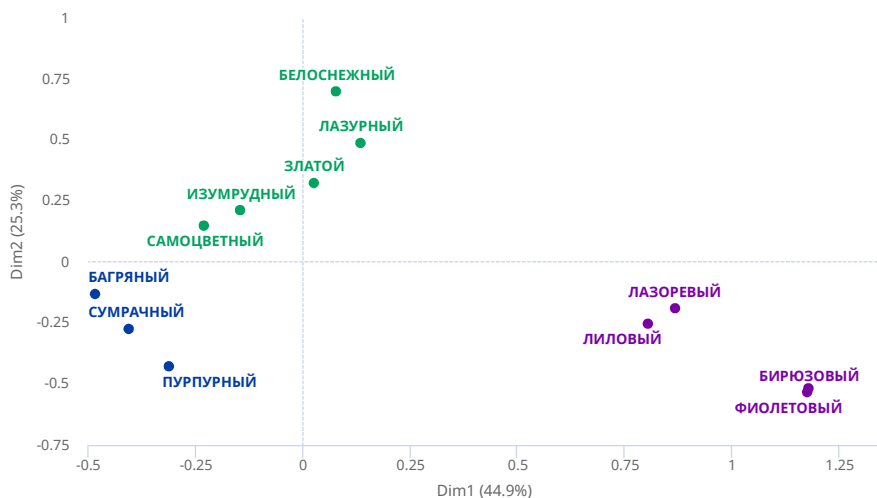


FIG. 2: Correspondence analysis plot: adjectives of colour (distances are defined by the authors' use profiles)

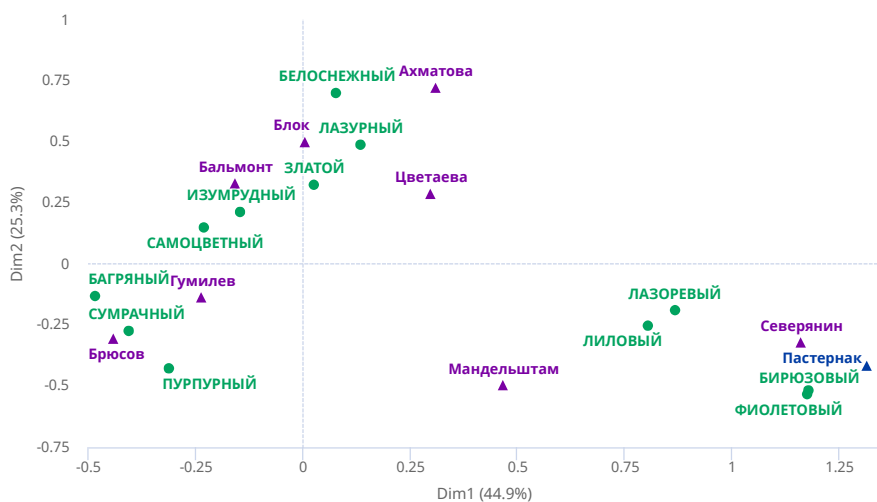


FIG. 3: Correspondence analysis plot: authors and adjectives of colour.
A supplementary point: Pasternak

лазурный 'azure' (the noun лазурь 'azure', denoting a gem, has two adjective derivative in Russian: лазурный 'azure' and later лазоревый 'azure').

The latter adjective, лазурный 'azure', alongside with златой 'golden', are a distinct feature of Aleksandr Blok and Konstantin Balmont. Meanwhile, the lexemes златой 'golden' and изумрудный 'emerald' belong to the poetic lexicon of Marina Tsvetaeva. The adjective самоцветный 'semiprecious' is included on the scale, so that it can be

attributed to the lexicon of Aleksandr Blok and Konstantin Balmont. However, this adjective is not a prominent feature of their poems.

The CA method shows that certain poets prefer different colours and hues. In some cases, their preferences are very explicit (see *лиловый* 'lilac' in Pasternak's poetry). Meanwhile, some poets turn out to be neutral with respect to the use of the given set of colour adjectives (as is the case with Nikolay Gumilev). It is notable that poets from the same poetic group do not necessarily favour the same colour hues. For example, although their aesthetic framework is the same, the acmeists Anna Akhmatova, Nikolay Gumilev, and Osip Mandel'shtam are rather distant from each other on the graph. Conversely, the subcorpora of authors belonging to different poetic groups can reveal a similar distribution of colour adjectives. This data supports the conclusion that, despite the internal influences within poetic groups, poets' colour preferences can be very different. However, this analysis only involves lexemes with considerable moderate frequencies. As a result, the statistical validity of our observation needs to be proven with additional tests.

3 The hue adjectives across meters

The following case study explores the associations between colour hue adjectives and the meters of texts. In Russian, most adjectives referring to non-banal colour hues are suffixed derivatives from words denoting material, such as gems: compare the noun *изумруд* 'emerald' (three syllables) and the adjective *изумрудный* 'emerald' (four syllables). For poetic texts, the syllabic length of a word is significant, as words are compelled to fit the metrical scheme of a verse line. Our analysis is focused on the multi-syllabic colour hue adjectives that we examined before. We excluded *златой* as the only 2/3-syllable word, and *сумрачный* 'murky', because its semantics is more vague than that of other colour hue words.

TAB. 1 reports the occurrences of the selected colour hue adjectives in texts containing the nine most frequent metrical types attested in our data: three-foot anapaests (An3), three- and four-foot amphibrachs (Aph3, Aph4), four- and five-foot trochees (Tr4, Tr5), three- to six-foot iambs (Ia3, Ia4, Ia5, Ia6).

TAB. 2 presents the same data via chi-grams, a metric of association which is calculated as the observed minus the expected occurrences divided by the square root of the expected occurrences. This metric underlies the use of the chi-squared statistic, which is basically the sum of the squares of the chi-grams. The values deviating the most from what is expected are highlighted. The table shows a strong association between multi-syllabic adjectives and certain metrical types. Adjectives that contain two syllables preceding the stressed one are, as expected, associated with anapaests and trochees. Adjectives with one syllable before the stress are associated with amphibrachs and neutral in regard to iambs. While this is the general tendency, a number of individual patterns can be observed. First, *лазоревый* 'azure' does not demonstrate any strong metrical associations. Second, the distribution of *изумрудный* 'emerald' is opposite that of the other words with the same syllable-stress structure.

Metric type	An3	Aph3	Aph4	Tr4	Tr5	Ia3	Ia4	Ia5	Ia6	Total
белоснежный 'snow-white'	23	6	3	46	9	7	40	33	4	171
бирюзовый 'turquoise'	7	5	2	15	3	1	23	10		66
изумрудный 'emerald'	15	6	8	29	11	6	46	25	11	157
фиолетовый 'violet'	6	1	1	13	5	1	14	7	5	53
лазоре́вый 'azure'	8	2	3	35	13	10	63	25	26	185
лазу́рный 'azure'	27	24	26	60	15	14	155	88	76	485
лило́вый 'lilac'	16	21	18	28	13	23	94	74	28	315
пурпу́рный 'tyrian purple'	10	8	3	19	9	2	78	48	38	215
багря́ный 'blood-red'	10	15	8	36	8	5	91	52	30	255
Total	122	88	72	281	86	69	604	362	218	1902

TAB. 1: Distribution of the color hue adjectives across metrical types

	An3	Aph3	Aph4	Tr4	Tr5	Ia3	Ia4	Ia5	Ia6
	-!-!-!	-!-!-!	-!-!-!-	!-!-!-	!-!-!-!	-!-!-	-!-!-!	-!-!-!-	-!-!-!-
белоснежный 'snow-white'	5,4	0,1	- 1,0	4,6	0,5	- 0,1	- 2,7	0,6	- 3,9
бирюзовый 'turquoise'	2,3	1,8	- 0,1	1,9	0,0	- 1,1	- 0,1	- 0,4	- 2,9
изумрудный 'emerald'	- 4,7	- 3,1	- 1,6	0,1	- 2,5	2,9	4,9	- 4,4	1,8
фиолетовый 'violet'	2,9	0,3	1,4	1,5	1,5	- 0,3	- 1,4	- 0,4	- 2,1
лазоре́вый 'azure'	- 0,2	- 1,7	- 1,2	1,8	1,6	0,7	- 0,4	- 1,2	0,3
лазу́рный 'azure'	1,0	1,8	2,8	- 0,9	- 1,4	- 1,5	- 1,4	0,5	1,5
лило́вый 'lilac'	0,4	3,1	2,6	- 2,4	- 0,3	2,6	- 1,8	2,7	- 2,1
пурпу́рный 'tyrian purple'	0,1	0,2	- 1,4	- 2,0	- 0,2	- 2,4	0,1	1,8	1,8
багря́ный 'blood-red'	- 1,7	- 1,7	- 1,2	- 2,1	2,7	- 0,6	- 0,1	1,6	1,7

TAB. 2: The chi-gram association of the colour hue adjectives and metrical types

One might wonder if the distribution of the colour hue words is fully powered by the syllable-stress profile of the word. TAB. 3 shows three stress patterns which occur in the Nominative singular masculine and the analogous forms (NomSgMasc-like, cf. бе-ло-сне!жны-й, бе-ло-сне!жны-х), the Genitive singular masculine and the analogous forms (GenSgMasc-like, cf. бе-ло-сне!жно-го-, бе-ло-сне!жны-ми-), and the short masculine form (ShortMasc, cf. бе-ло-сне!же-н). The stress patterns are coded according to standard (prosaic) Russian. There are two major groups of adjectives: those that have two unstressed syllables before the stressed syllable, and those that have one unstressed syllable before the stressed one. If we exclude from consideration the short masculine form, which occurs very rarely in our data, we can see that each major group has two subtypes, фиолетовый and лазоре́вый being one syllable longer than the other words in their groups.

TAB. 4 illustrates the association between stress patterns and metrical types calculated with regard to a sample of all stressed words (from texts created in the period 1900-1917). As one can see, adjectives that have two syllables before or after the stressed one in NomSgMasc are distributed similarly across metrical types, in contrast to the adjectives with the syllable-stress profile “- ! -”. This can explain the differences in the behaviour of лазоре́вый and лазу́рный, two adjectives that refer to the same colour hue ('azure') but have different suffixes and thus different syllable-

	NomSgMasc-like	GenSgMasc-like	ShortMasc
белоснежный ‘snow-white’	--!-	--!--	--!-
бирюзовый ‘turquoise’	--!-	--!--	--!
изумрудный ‘emerald’	--!-	--!--	--!-
фиолетовый ‘violet’	--!--	--!---	--!-
лазоре́вый ‘azure’	-!--	-!---	-!-
лазу́рный ‘azure’	-!-	-!--	-!-
лило́вый ‘lilac’	-!-	-!--	-!
пурпу́рный ‘tyrian purple’	-!-	-!--	-!-
багря́ный ‘blood-red’	-!-	-!--	-!

TAB. 3: The syllable-stress profiles of the color hue adjectives

	An3	Aph3	Aph4	Tr4	Tr5	Ia3	Ia4	Ia5	Ia6	Total
	-!-!-	-!-!-	-!-!-!-	!-!-!-	!-!-!-!-	-!-!-	-!-!-!-	-!-!-!-	-!-!-!-	
--!-	45,00	45,30	37,63	31,94	22,11	-11,14	-38,96	-27,77	-19,05	3911
--!--	28,15	20,44	11,85	24,57	13,15	-5,99	-20,99	-14,98	-10,25	1132
-!--	45,11	25,68	19,30	31,62	6,14	18,35	-29,01	-31,44	-15,76	5728
-!-	-27,09	-20,63	-15,90	-19,69	-8,64	-1,85	20,15	17,44	10,30	70196
Total	9301	5762	4139	3506	634	2568	31500	16048	7511	80969

TAB. 4: The chi-gram association of four multi-syllabic stress patterns and metric types

stress profiles. Due to the intermediate status of лазоре́вый (more than three syllables in NomSgMasc, but similarity to the scheme “- ! -”), this adjective does not demonstrate any strong associations, see TAB. 2.

Although the general tendency mentioned above explains a lot about the distribution of adjectives across metrical types, one can see that adjectives with the same syllable-stress profile actually demonstrate more diverse behaviours. The prominent exception is изумрудный ‘emerald’, which we discussed above. Other explanatory factors can be suggested, especially the position of the word in the line and the phonetic characteristics of the first syllable. In sum, further qualitative analysis of particular contexts has to be carried out to study the underlying characteristics of the word use.

4 Conclusion

This paper presented two explanatory case studies on colour hue adjectives with a focus on quantitative metrics extracted via a new frequency database comprising more than 13 million tokens with several layers of linguistic, versological, and meta-textual annotation.

The study aimed to explain chronological changes, as well as similarities and differences in poetic vocabulary, with respect to quantitative metrics, semantics, syllabic structure and accentual patterns of a word.

We explored the use of colour adjectives in the Silver Age, applying the method of Correspondence Analysis, which offers visualization of multidimensional frequency

associations of lexemes and authors. This method supports contrastive stylistic analysis and identifies similarities between different poets. In this case, we defined subcorpora of authors traditionally seen as key figures of the Silver Age such as Aleksandr Blok, Konstantin Balmont, Anna Akhmatova, Nikolay Gumilev, and Marina Tsvetaeva. This study has revealed the stylistic differences in individual poetic lexicons and demonstrated that despite the internal influences within poetic schools, poets' colour preferences can be very different. Thus, by employing the annotation of the Russian Poetry corpus, one can include analysis of semantic layers in diachronic and comparative research.

We analysed the syllabic structure of the most prominent colour hue adjectives of the Silver Age and classified them into four groups according to their stress-syllable type and the number of unstressed syllables before and after the stressed syllable. The distribution of the selected colour hue adjectives across metrical types highlights strong associations of multi-syllabic adjectives with certain meters, although some words have a more complex distribution.

Acknowledgment

The research was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2018 (grant #18-05-0047) and by the Russian Academic Excellence Project "5-100".

References

- Divjak, D. – Sharoff, S. – Erjavec, T. (2017). *Slavic Corpus and Computational Linguistics*. *Journal of Slavic Linguistics* 25(2).
- Dresher, B. – Friedberg, N. (2008). *Formal Approaches to Poetry. Recent Developments in Metrics*. Berlin, Boston: De Gruyter Mouton.
- Gasparov, M. (2005). *Sovremennyy russkij stikh: metrika i ritmika* [Contemporary Russian Verse: Metrics and Rhythmics]. Moscow: Nauka.
- Grishina, E. – Korchagin, K. – Plungian, V. – Sichinava, D. (2009). Poeticheskij korpus v ramkakh Natsional'nogo korpusa russkogo yazyka: obshchaya struktura i perspektivy ispol'zovaniya [The corpus of poetry within the Russian National Corpus: a general outline and perspectives of use]. In *Natsional'nyj korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy* [The Russian National Corpus: 2006–2008. New results and prospects]. St. Petersburg: Nestor-Istoriya Publ.
- Jacobs, A. M. (2018). The Gutenberg English Poetry Corpus: Exemplary Quantitative Narrative Analyses. *Frontiers in Digital Humanities* (5). Retrieved from <https://www.frontiersin.org/articles/10.3389/fdigh.2018.00005/full>
- Jakobson, R. (1973). *Questions de poétique*. Paris: Seuil.

- Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. STHDA.
- Kizhner, I. – Terras, M. – Manovich, L. – Orekhov, B. – Bonch-Osmolovskaya, A. – Rumyantsev, M. (2018). The history and context of the Digital Humanities in Russia. In *Proceedings of DH 2018*, Mexico City, June 26–29, 2018, 416–419.
- Kopotev, M. – Lyashevskaya, O. – Mustajoki, A. (2018). *Quantitative Approaches to the Russian Language*. Routledge.
- Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.
- Lyashevskaya, O. – Vlasova, E. – Litvintseva, K. – Sechina, E. (2018). Data Analysis Tool for the Corpus of Russian Poetry / *NRU HSE. Series WP BRP "Linguistics"*. No. 77. Retrieved from <https://wp.hse.ru/data/2018/12/29/1143017363/77LNG2018.pdf>
- Russian National Corpus. Available at <http://ruscorpora.ru/en/index.html>
- Taranovsky, K. F. (2010). *Russkie dvuslozhnye razmery. Stati o stixe*. [Russian binary meters. Articles about verse]. Bailey, J. (Ed.). Moscow: Yazyki slav. kultur.
- Yarkho, B. I. (2006). *Metodologiya tochnogo literaturovedeniya: Izbrannye trudy po teorii literatury* [Methodology of literary studies: Selected works in the theory of literature]. Shapir, M.I. (Ed.). Moscow: Yazyki slav. kultur.