SEMANTIC HALO OF A METER: A KEYWORD-BASED APPROACH

Piperski A. Ch. (apiperski@gmail.com)

Russian State University for the Humanities; National Research University Higher School of Economics, Moscow, Russia

The semantic halo of a meter (*semantičeskij oreol metra*) is a notion that was introduced by Mikhail Gasparov to describe semantic invariants of poetic texts composed using the same metrical scheme. Most studies addressing this phenomenon have been based on expert knowledge of the text corpus. In this study, I propose an automated approach to analyzing the semantic halo of various meters based on keyword extraction, using a simple measure of keyness as developed by Adam Kilgarriff. The method is applied to texts from the Poetic subcorpus of the Russian National Corpus. It allows us to discern basic motifs that are very close to those identified by literary scholars, which proves that it is a promising way to analyze the semantic halo of a meter. Some novel associations can also be inferred from the keyword lists. Clearly, keyword extraction cannot replace profound expert knowledge, but it can serve as a useful first step of the analysis.

Keywords: semantic halo of a meter, verse, keyword extraction, Russian National Corpus

СЕМАНТИЧЕСКИЙ ОРЕОЛ МЕТРА И ИЗВЛЕЧЕНИЕ КЛЮЧЕВЫХ СЛОВ

Пиперски А. Ч. (apiperski@gmail.com)

РГГУ; НИУ ВШЭ, Москва, Россия

Понятие семантического ореола метра было введено М. Л. Гаспаровым для описания семантических инвариантов поэтических текстов, написанных по одной и той же метрической схеме. Большинство исследований в этой области базируются на экспертном знании текстов, однако в данной работе предлагается автоматизированный подход к анализу семантических ореолов различных метров. Он опирается на извлечение ключевых слов методом, предложенным А. Килгарриффом. Этот подход иллюстрируется на материале текстов из Поэтического

корпуса Национального корпуса русского языка. Он позволяет выявить базовые мотивы, которые близки к мотивам, выделяемым для соответствующих метрических схем литературоведами, что демонстрирует перспективность описываемого метода. Кроме того, он позволяет обнаружить некоторые ранее не отмечавшиеся ассоциации. Разумеется, извлечение ключевых слов не может заменить собой глубокого экспертного знания, однако может служить в качестве отправной точки для последующего анализа.

Ключевые слова: семантический ореол метра, стих, извлечение ключевых слов, Национальный корпус русского языка

1. Introduction

The semantic halo of a meter (*semantičeskij oreol metra* in Russian) is a notion that was introduced by Gasparov (1979) in order to describe semantic invariants of poetic texts composed using the same meter. Most studies addressing this phenomenon have been based on expert knowledge of the text corpus, the most prominent among them being the studies by Gasparov himself (1999). This book contains a collection of essays about different metrical schemes, including trochaic tetrameter with alternating dactylic and masculine endings (Х4дмдм = T4dmdm), iambic trimeter (Я3 = I3), dactylic tetrameter with dactylic endings (Д4дддд = D4dddd), trochaic pentameter (Х5 = T5), etc. All of these metrical schemes are shown to have some culturally rooted semantic associations, and the evolution of these associations can also be traced historically. For instance, Gasparov establishes five principal motifs present in the Russian trochaic pentameter, namely, Way, Night, Love, Landscape, and Life & Death (Gasparov 1999: 331–370). They are easily discernible in Lermontov's "Vyxožu odin ja na dorogu" ("Alone I step out onto the road", 1841), which is probably the best-known and the most influential poem in this meter, and can be encountered in various combinations in other poems employing trochaic pentameter. Yet another collection of analyses made along similar lines was published by Wachtel (1998). However, it is clear that the availability of large amounts of digitized poetical texts calls for an automated approach to this problem (Birnbaum & Thorsen 2015).

2. Methodology and data

The main hypothesis underlying the present study is that the semantic halo of a meter can be at least hinted at by the keywords used in the poems written in this meter. For instance, Lermontov's "Vyxožu odin ja na dorogu" directly contains the words *doroga* 'way, path', *noč'* 'night', *ljubov'* 'love', *žizn'* 'life', which correspond to four out of five motifs identified by Gasparov, as well as many words pertaining to Landscape, including *tuman* 'fog', *zvezda* 'star', *nebesa* 'heaven', *zemlja* 'earth', and *dub* 'oak'. If some of these motifs are repeatedly occurring in poems in trochaic pentameter and are more frequent here than in poems using other metrical schemes, we might expect that the corresponding words will occur more frequently in trochaic pentameter—in other words, they will be the keywords of the iambic pentameter corpus

as compared to the corpus consisting of poems written in other meters. These key-words can be identified using some keyness measure. Some scholars have expressed a skeptical attitude towards equating the semantic halo of a meter with its keywords (Brogan 2012: 1264), but it must be noted that keywords can be useful as a starting point for a further detailed analysis, which I aim to show.

In order to study semantic invariants of texts with different metrical schemes, I use the Poetic subcorpus of the Russian National Corpus (RNC), which contains a large collection of Russian poetry from the 18th century up to the 2nd half of the 20th century. It includes 78868 metrically annotated poetic texts, their total size amount-ing to 11 million words. For the purposes of this study, only texts containing no more than 500 words were included in the sample, since longer texts, such as Pushkin's "Eu-gene Onegin" or Nekrasov's "Komu na Rusi žit' xorošo" ("Who is Happy in Russia?"), are hardly ever homogeneous and can barely serve as good representatives of the semantic halo of a meter. Moreover, some recurring words and names from longer texts, e.g., *Tat'jana* 'Tatyana (proper name)' in "Eugene Onegin", make frequency lists skewed and are likely to be identified as keywords typical of the metrical scheme rather than of individual texts. This restriction led to the exclusion of some 3% of the texts, leaving a corpus of 76,559 poems, containing 8 million tokens (including punc-tuation marks which are treated as separate tokens). These texts were lemmatized and automatically disambiguated with Mystem (pymystem3 library for Python). Clearly, this leads to some inaccuracies, but these are unavoidable and cannot be corrected manually because of the large corpus size.

There are many possible ways of identifying keywords in corpus linguistics (see, e.g., Scott 1997; Bondi & Scott (eds.) 2010; Culperer & Demmen 2015). For the pur-pose of this study, I use the Add-N keyness measure proposed by Kilgarriff (2009). Once the frequencies in the focus corpus (fc) and the frequencies in the reference corpus (rc) are expressed in ipm, the keyness measure K for a word w is calculated as follows based on the frequencies (f) of this word in the two corpora:

$$K = (f_{fc}(w) + N) \, / \, (f_{rc}(w) + N).$$

Depending on the choice of N, this measure favors different types of keywords. If N is low (e.g., $N = 1$), it selects low-frequency words, their keyness being large in terms of effect size but not necessarily statistically significant; if N is high (e.g., $N = 1,000$), it selects high-frequency words, their keyness being not so large in terms of effect size but much more statistically significant. The value of N best suited for the purpose of a study has to be identified by the researcher. In the present case, $N = 3,000$ was chosen in order to extract only high-frequent keywords.

3.   Results and discussion

There are 20 metrical schemes represented by more than 50,000 tokens. For these schemes, keywords with $K > 1.1$ were selected. The results are listed in the Appendix.

It is obvious that there are some metrical schemes that do not possess an identifi-able semantic halo. One of them is (1) Iambic tetrameter with alternating feminine and masculine endings (I4fm = Я4жм), which also happens to be the most popular

metrical scheme overall. It is characterized by the iambic words *kogdá* 'when', *dušá* 'soul', and *mečtá* 'dream', whose high frequency can be explained by meter: a subordinating conjunction like *kogdá* 'when' is expected to appear at the beginning of a verse since syntactic boundaries are likely to coincide with verse boundaries (enjambements do of course occur, but they are rare), and iambic conjunctions are favored here because they fit the meter. The same is true of *dušá* 'soul' and *mečtá* 'dream', which are favored not only at the beginning of a verse, but also verse-internally after possessive pronouns denoting a singular possessor (*mojá / tvojá / egó / ejó dušá* 'my / your / his / her soul'). A similar preference for words for which prosodic template is identical to the verse foot is visible in the lists for all metrical schemes. For example, the lists for trochaic schemes in (2) and (9) contain trochaic words *mílyj* 'dear', *sérdce* 'heart', *pésnja* 'song', *étot* 'this', *ésli* 'if', *tól'ko* 'only', etc. close to the top. The amphibrachic scheme (10) favors the words *Rossíja* 'Russia', *xolódnyj* 'cold', and *dalĕ́kij* 'far-away', whereas the list for anapaest in (11) includes anapaestic words *molodój* 'young', *golubój* 'light blue', and *zolotój* 'golden'.

Metrical scheme (2) Trochaic tetrameter with feminine and masculine endings (T4fm = Х4жм) is linked to Love and Song, which is exemplified by the keywords *mílyj* 'dear', *serdce* 'heart', and *pesnja* 'song'. It also features emotional language and dialogs, which is reflected in the high frequency of quotation marks («) and exclamation marks (*!*).

Metrical scheme (7) Iambic hexameter with alternating feminine and masculine endings (I6fm = Я6жм) has obvious romantic associations. Its keywords include such nouns as *duša* 'soul', *ljubov'* 'love', *luč* 'ray', *krasota* 'beauty', *nadežda* 'hope'. A similar conclusion holds true for (3) Iambic tetrameter with various patterns of feminine and masculine endings (I4f,m = Я4ж,м), which is further linked to the topic of creativity through *pevec* 'singer', *muza* 'Muse', and *vdoxnovenie* 'inspiration'.

Metrical scheme (9) Trochaic pentameter with alternating feminine and masculine endings (T5fm = Х5жм) is an interesting test case for the method presented here, since it is was thoroughly studied by Gasparov (see Section 1 above). Four out of five motifs listed by Gasparov are also observable in the keyword list, namely Way (*uxodit'* 'go away', *put'* 'way'), Life & Death (*žit'* 'to live'), Night (*noč'* 'night', *zvezda* 'star'), and Landscape (*zemlja* 'earth', *zvezda* 'star'). This shows that the keyword approach yields results similar to an in-depth analysis by a literary scholar.

Another case where keywords conform to Gasparov's analysis is in metrical scheme (10) Amphibrachic trimeter with alternating feminine and masculine endings (Ab3fm = Аф3жм). One of its possible semantic associations is "elevated intonation" (Gasparov 1999: 194–202), which is closely linked to patriotism. In the keyword list, this is expressed by the high frequency of the words *Rossija* 'Russia' and *russkij* 'Russian'. A closer look at the texts shows that these words are especially characteristic of the poems in this meter composed by Andrei Bely, Georgy Ivanov, and Yaroslav Smelyakov.

Metrical scheme (11) Anapaestic trimeter with alternating feminine and masculine endings (An3fm = Ан3жм) seems to be closely connected to the motif of Landscape (*zemlja* 'ground, earth', *sneg* 'snow', *prostor* 'open space', *vperedi* 'in front').

Metrical scheme (12) *Dolnik* with alternating feminine and masculine endings (Dk3fm = Дк3жм) is a form that has entered Russian poetry relatively recently, and

most of the poems that represent it come from the 20th century. It shows a turn to more personal topics, which is exemplified by the prevalence of body part names (*glaz* 'eye', *ruka* 'hand / arm'). It also

Metrical scheme (15) Iambic trimeter with alternating feminine and masculine endings (I3fm = Я3жм) is a metrical scheme associated with the motif of Love, which is confirmed by the fact that its keywords include words like *milyj* 'dear', *ljubit'* 'to love', *ljubov'* 'love', *ax* 'o!', *vesna* 'spring'.

Metrical scheme (18) Trochaic tetrameter with feminine endings (T4f = Х4ж) was quite popular in the middle of the 19th century and is often used for describing nature: *solnce* 'sun', *more* 'sea', *tuča* 'cloud', *nebo* 'sky', *les* 'forest'.

These illustrative analyses of the keyword lists show clearly that keywords can be linked to the notion of the semantic halo of a meter. Of course, one must proceed with caution, since general keyword lists for a metrical scheme do not take into account any historical development and may include keywords that were relevant in different time periods. Some of the items may be invited by the meter formally (cf. discussion of *kogda*) rather by a cultural semantic association, which also must be differentiated from the semantic halo.[1] One should also propose an operationalizable procedure for identifying semantic fields based on keywords, since the interpretations given above are often fairly impressionistic. However, keyword lists can definitely serve as a starting point for more detailed investigations.

4.   Conclusion

The automated extraction of keywords from large poetic corpora has been shown to be a useful way for identifying the semantic halo of various meters. It cannot replace a detailed analysis of texts performed by a literary scholar, but this approach can facilitate the primary identification of relevant motifs and serve as a starting point for further analysis.

References

1. *Birnbaum D., Thorsen E.* (2015), Enabling the automated identification and analysis of meter and rhyme in Russian verse, Paper presented at Computational approaches to poetry: a day of workshops, University of Helsinki, Finland, June 2015.
2. *Bondi M., Scott M.* (eds., 2010), Keyness in texts, John Benjamins, Amsterdam; Philadelphia.
3. *Brogan T. V. F.* (2012), Trochaic, In Greene R. (ed.), The Princeton encyclopedia of poetry and poetics, 4th ed., Princeton University Press, Princeton, pp. 1462–1463.
4. *Culperer J., Demmen J.* (2015), Keywords, In Biber D., Reppen R. (eds.), The Cambridge handbook of English corpus linguistics, Cambridge University Press, Cambridge.

---

[1]   This was pointed out by an anonymous reviewer to whom I am very grateful.

5. *Gasparov M. L.* (1979), Semantic halo of a meter: Towards the semantics of the Russian iambic trimester [Semantičeskij oreol metra: k semantike russkogo trëhstopnogo jamba], In Linguistics and poetics [Lingvistika i poètika], Nauka, Moscow, pp. 282–308.
6. *Gasparov M. L.* (1999), Meter and semantics: On one of the mechanisms of cultural memory [Metr i smysl: ob odnom iz mehanizmov kul'turnoj pamjati], RGGU, Moscow.
7. *Kilgarriff A.* (2009), Simple maths for keywords, In Proceedings of Corpus Linguistics Conference CL2009, University of Liverpool, UK, July 2009.
8. RNC. Russian National Corpus. www.ruscorpora.ru
9. *Scott M.* (1997), PC analysis of key words — and key key words, System 25(2), pp. 233–245.
10. *Wachtel M.* (1998), The development of Russian verse: Meter and its meanings, Cambridge University Press, Cambridge.

Appendix

Word	Focus corpus freq.	Reference corpus freq.	K
1. Iambic tetrameter with alternating feminine and masculine endings (I4fm = Я4жм; 695,663 tokens)			
твой	4,914.7	3,433.7	1.230
когда	3,017.3	1,975.3	1.209
но	6,858.2	5,507.9	1.159
свой	4,216.1	3,241.2	1.156
мир	2,396.3	1,751.8	1.136
мечта	1,085.3	671.3	1.113
душа	3,018.7	2,427.9	1.109
2. Trochaic tetrameter with feminine and masculine endings (T4fm = Х4жм; 331,449 tokens)			
милый	1,777.0	940.6	1.212
!	4,634.2	3,312.7	1.209
сердце	3,330.8	2,446.0	1.162
«	1,834.4	1,196.8	1.152
песня	1,653.3	1,071.5	1.143
ж	1,412.0	873.1	1.139
наш	2,380.5	1,726.5	1.138
ходить	775.4	358.7	1.124
ах	956.4	522.7	1.123
уж	1,390.9	942.2	1.114
свет	2,139.1	1,616.2	1.113
словно	902.1	516.0	1.110

Word	Focus corpus freq.	Reference corpus freq.	K
ль	1,173.6	772.8	1.106
око	992.6	612.1	1.105
волна	1,249.1	857.1	1.102
тихо	859.9	507.5	1.100
3. Iambic tetrameter with various patterns of feminine and masculine endings (I4f,m = Я4ж,м; 292,724 tokens)			
когда	3,279.5	2,019.7	1.251
вы	4,195.1	2,860.6	1.228
мой	8,007.5	5,962.2	1.228
друг	2,551.9	1,540.9	1.223
поэт	1,298.2	610.7	1.190
мечта	1,369.9	682.1	1.187
душа	3,388.9	2,444.7	1.173
она	5,298.5	4,072.3	1.173
ваш	1,164.9	590.6	1.160
любовь	2,473.3	1,727.4	1.158
стих	1,311.8	733.5	1.155
твой	4,516.2	3,526.1	1.152
свой	4,249.7	3,290.8	1.152
сей	1,028.3	514.3	1.146
муза	683.2	244.3	1.135
взор	1,246.9	743.6	1.134
милый	1,472.4	956.3	1.130
певец	591.0	198.5	1.123

Word	Focus corpus freq.	Reference corpus freq.	K
но	6,603.5	5,588.0	1.118
жизнь	2,695.4	2,108.0	1.115
око	1,021.4	612.9	1.113
надежда	744.7	371.3	1.111
сердце	3,064.3	2,460.5	1.111
природа	761.8	391.3	1.109
слава	949.7	563.2	1.108
вдохновение	440.7	108.7	1.107
пора	1,195.7	789.5	1.107
живой	1,168.3	764.3	1.107
ж	1,287.9	880.5	1.105
ее	1,117.1	729.1	1.104
ум	806.2	448.0	1.104
его	1,612.4	1,186.7	1.102

4. Iambic pentameter with alternating feminine and masculine endings (I5fm = Я5жм; 291,540 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
но	7,011.0	5,572.8	1.168
сонет	548.8	64.2	1.158
мечта	1,238.3	687.2	1.149
мир	2,483.4	1,782.2	1.147
сон	2,112.9	1,494.2	1.138
когда	2,733.8	2,040.5	1.138
душа	3,203.7	2,451.8	1.138
день	3,430.1	2,672.8	1.133
и	50,970.7	44,841.9	1.128
как	11,970.9	10,410.7	1.116
о	5,049.0	4,218.5	1.115
любовь	2,233.0	1,736.6	1.105

5. Iambic tetrameter with alternating masculine and feminine endings (I4mf = Я4мж; 219,314 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
твой	4,582.5	3,533.5	1.161
но	6,917.0	5,588.7	1.155
когда	2,799.6	2,045.1	1.150
.	7,477.9	6,142.5	1.146
день	3,469.9	2,678.7	1.139
душа	3,082.3	2,462.2	1.114
тоска	1,108.0	707.1	1.108
иной	797.9	440.6	1.104

6. Iambic pentameter with alternating masculine and feminine endings (I5mf = Я5мж; 173,737 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
но	7,897.0	5,574.7	1.271
сонет	627.4	69.7	1.182
как	12,329.0	10,426.2	1.142
с	460.5	38.6	1.139
.	7,315.7	6,153.9	1.127
чей	719.5	305.0	1.125
вновь	949.7	548.0	1.113
мир	2,336.9	1,796.0	1.113
когда	2,618.9	2,053.5	1.112
тень	1,421.7	988.0	1.109
сон	1,991.5	1,506.2	1.108
миг	794.3	441.2	1.103

7. Iambic hexameter with alternating feminine and masculine endings (I6fm = Я6жм; 144,418 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
душа	3,967.6	2,451.9	1.278
когда	3,365.2	2,041.9	1.262
я	26,679.5	20,894.4	1.242
твой	5,061.7	3,534.7	1.234
сонет	782.5	68.9	1.232
любовь	2,783.6	1,735.7	1.221
ты	13,114.7	10,510.3	1.193
себя	2,264.3	1,461.1	1.180
мечта	1,357.2	695.3	1.179
свой	4,417.7	3,305.8	1.176
толпа	948.6	403.1	1.160
их	1,620.3	995.4	1.156
о	5,352.5	4,228.5	1.155
луч	1,225.6	657.5	1.155
вновь	1,073.3	547.3	1.148
но	6,827.4	5,603.0	1.142
любить	1,973.4	1,362.3	1.140
мой	7,277.5	6,014.1	1.140
нет	2,797.4	2,110.2	1.134
красота	865.5	410.1	1.134
то	4,016.1	3,189.8	1.133
миг	893.2	440.7	1.132

Word	Focus corpus freq.	Reference corpus freq.	K
сомнение	512.4	125.7	1.124
и	50,845.5	44,958.7	1.123
лишь	1,744.9	1,224.4	1.123
сон	2,049.6	1,506.9	1.120
надежда	768.6	377.9	1.116
мир	2,354.3	1,797.7	1.116
ни	2,735.1	2,156.7	1.112
тишина	920.9	526.2	1.112
жизнь	2,686.6	2,119.2	1.111
здесь	1,620.3	1,166.4	1.109
пред	983.3	591.7	1.109
который	1,211.8	802.6	1.108
как	11,861.4	10,441.9	1.106
тьма	865.5	498.4	1.105
природа	754.8	398.4	1.105
путь	1,405.6	998.3	1.102

8. Iambic pentameter with various patterns of feminine and masculine endings (I4f,m = Я5ж,м; 134,355 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
и	55,070.5	44,894.2	1.212
но	7,256.9	5,597.2	1.193
я	24,561.8	20,937.8	1.151
он	9,214.4	7,925.5	1.118
его	1,667.2	1,194.3	1.113
ни	2,701.8	2,158.0	1.105
когда	2,575.3	2,057.0	1.102

9. Trochaic pentameter with alternating feminine and masculine endings (T5fm = Х5жм; 125,119 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
этот	4,459.8	2,694.8	1.310
если	1,726.4	1,017.1	1.177
только	2,509.6	1,679.6	1.177
земля	2,693.4	1,872.6	1.168
словно	1,095.0	523.0	1.162
год	2,134.0	1,447.9	1.154
ночь	2,837.3	2,058.9	1.154
мочь	2,349.8	1,650.9	1.150
такой	1,598.5	1,009.1	1.147
чтобы	959.1	472.6	1.140
что	9,295.2	7,807.7	1.138
сколько	743.3	288.5	1.138

Word	Focus corpus freq.	Reference corpus freq.	K
над	3,316.8	2,564.3	1.135
звезда	1,582.5	1,049.5	1.132
город	1,095.0	623.9	1.130
уходить	1,158.9	682.5	1.129
синий	967.1	513.1	1.129
ветер	1,726.4	1,195.7	1.126
каждый	1,422.6	926.7	1.126
путь	1,486.6	998.0	1.122
золотой	1,031.0	608.4	1.117
голубой	839.2	450.5	1.113
по	4,419.8	3,679.1	1.111
темный	1,134.9	731.1	1.108
жить	1,830.3	1,378.0	1.103
разве	511.5	191.6	1.100

10. Amphibrachic trimeter with alternating feminine and masculine endings (Ab3fm = Аф3жм; 100,252 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
туман	1,216.9	522.0	1.197
и	53,594.9	44,956.7	1.180
над	3,551.1	2,563.7	1.177
идти	2,194.5	1,541.7	1.144
Россия	728.2	260.4	1.143
холодный	887.8	423.1	1.136
далекий	857.8	396.0	1.136
под	3,032.4	2,319.9	1.134
за	5,157.0	4,208.6	1.132
она	4,897.7	4,107.2	1.111
на	15,371.3	13,563.8	1.109
а	6,214.3	5,350.2	1.103
шуметь	638.4	300.1	1.102
русский	698.2	360.5	1.101

11. Anapaestic trimeter with alternating feminine and masculine endings (An3fm = Ан3жм; 90,658 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
этот	5,085.0	2,695.3	1.420
только	3,011.3	1,677.4	1.285
словно	1,456.0	521.4	1.265
над	3,915.8	2,560.7	1.244
это	2,603.2	1,689.0	1.195
молодой	1,070.0	492.0	1.166
голубой	1,014.8	450.2	1.164

Word	Focus corpus freq.	Reference corpus freq.	K
земля	2,669.4	1,876.4	1.163
ты	12,729.2	10,532.4	1.162
ни	2,945.1	2,158.2	1.153
золотой	1,158.2	608.7	1.152
чтобы	1,003.8	474.2	1.152
по	4,665.9	3,679.5	1.148
дорога	1,290.6	748.4	1.145
за	5,140.2	4,209.9	1.129
я	23,991.3	20,964.3	1.126
сколько	705.9	290.9	1.126
.	7,291.1	6,166.3	1.123
снег	1,125.1	725.8	1.107
простор	606.7	259.7	1.106
до	2,117.8	1,628.0	1.106
на	15,310.3	13,566.7	1.105
не	19,468.8	17,330.0	1.105
от	4,081.3	3,407.3	1.105
впереди	430.2	105.9	1.104
кто-то	606.7	272.9	1.102
12. *Dolnik* with alternating feminine and masculine endings (Dk3fm = Дк3жм; 86,034 tokens)			
это	2,812.8	1,687.3	1.240
только	2,615.2	1,682.5	1.199
.	7,787.6	6,161.6	1.177
ветер	1,917.8	1,196.2	1.172
до	2,417.6	1,625.0	1.171
этот	3,661.3	2,712.1	1.166
словно	1,034.5	526.5	1.144
глаз	2,068.9	1,433.2	1.143
наш	2,406.0	1,746.4	1.139
на	15,831.0	13,562.0	1.137
надо	1,022.9	602.6	1.117
дверь	790.4	410.0	1.112
последний	1,115.8	700.7	1.112
вечер	941.5	552.9	1.109
рука	2,510.6	1,975.2	1.108
по	4,382.0	3,683.1	1.105
море	1,383.2	971.9	1.104

Word	Focus corpus freq.	Reference corpus freq.	K
13. Iambic hexameter with paired feminine and masculine endings (I6ffmm = Я6жжмм; 77,957 tokens)			
сей	3,412.1	504.8	1.830
себя	3,194.1	1,458.7	1.389
то	5,541.5	3,181.8	1.382
ты	15,329.0	10,510.4	1.357
который	2,129.4	797.0	1.351
свой	5,413.2	3,305.3	1.334
коль	1,090.3	164.9	1.292
зреть	1,103.2	178.4	1.291
когда	3,412.1	2,052.5	1.269
иль	1,770.2	783.3	1.261
твой	5,079.7	3,547.3	1.234
стих	1,603.4	746.3	1.229
чтоб	2,321.8	1,330.6	1.229
сие	795.3	87.9	1.229
иметь	897.9	206.2	1.216
о	5,772.4	4,233.8	1.213
тот	3,976.6	2,766.2	1.210
он	10,159.4	7,925.4	1.204
во	2,424.4	1,508.8	1.203
дух	1,423.9	680.4	1.202
мой	7,824.8	6,019.3	1.200
сам	1,860.0	1,061.3	1.197
!	4,566.6	3,355.5	1.191
уж	1,693.2	953.5	1.187
всякий	872.3	268.3	1.185
ль	1,475.2	782.7	1.183
ныне	756.8	178.4	1.182
любезный	679.9	116.8	1.181
кто	3,117.1	2,203.6	1.176
мочь	2,462.9	1,654.0	1.174
кой	615.7	85.9	1.172
она	5,310.6	4,105.4	1.170
муза	795.3	255.1	1.166
уже	1,359.7	748.8	1.163
являть	564.4	67.1	1.162
я	24,821.4	20,961.0	1.161

Word	Focus corpus freq.	Reference corpus freq.	K
великий	949.2	411.6	1.158
ум	1,000.6	455.8	1.158
наука	590.1	118.1	1.151
слава	1,103.2	572.1	1.149
к	5,644.1	4,557.9	1.144
труд	936.4	449.4	1.141
око	1,128.8	622.9	1.140
взор	1,257.1	757.1	1.133
желать	590.1	174.0	1.131
они	4,605.1	3,731.8	1.130
утеха	461.8	72.3	1.127
другой	1,372.6	887.8	1.125
хотя	602.9	206.0	1.124
всегда	885.1	457.3	1.124
богиня	487.4	102.5	1.124
тогда	962.1	527.7	1.123
от	4,194.6	3,407.3	1.123
мысль	949.2	521.7	1.121
муж	564.4	180.7	1.121
дар	628.6	242.3	1.119
писать	679.9	288.0	1.119
быть	8,722.8	7,493.8	1.117
днесь	397.7	41.6	1.117
сатира	384.8	30.3	1.117
пред	1,013.4	594.7	1.116
б	1,180.1	745.3	1.116
однако	423.3	71.5	1.115
знать	2,309.0	1,766.5	1.114
приятный	436.1	84.9	1.114
глас	487.4	138.9	1.111
достойный	423.3	85.0	1.110
вид	602.9	246.9	1.110
красота	782.5	414.8	1.108
котор	346.3	22.7	1.107
оставлять	667.0	313.0	1.107
славный	474.6	137.5	1.107
какой	1,257.1	850.2	1.106
день	3,296.7	2,694.5	1.106
прекрасный	821.0	455.4	1.106
российский	359.2	47.6	1.102
царь	744.0	396.5	1.102

Word	Focus corpus freq.	Reference corpus freq.	K
честь	551.6	226.0	1.101
ко	756.8	413.7	1.100
14. Trochaic tetrameter with various patterns of feminine and masculine endings (T4f,m = Х4ж,м; 72,217 tokens)			
дева	1,135.5	294.9	1.255
конь	1,218.5	470.9	1.215
милый	1,758.6	968.1	1.199
это	2,534.0	1,691.8	1.180
если	1,661.7	1,022.4	1.159
или	1,343.2	762.6	1.154
очень	706.2	216.2	1.152
марья	429.3	9.6	1.139
вода	1,620.1	1,057.4	1.139
тихо	969.3	518.0	1.128
ну	803.1	387.4	1.123
из	3,988.0	3,241.4	1.120
ходить	775.4	372.3	1.120
нет	2,714.0	2,117.2	1.117
волна	1,315.5	869.3	1.115
бить	706.2	324.7	1.115
река	1,066.2	661.9	1.110
дед	443.1	106.8	1.108
пить	775.4	411.1	1.107
речка	415.4	85.8	1.107
грудь	1,273.9	865.1	1.106
у	3,586.4	2,963.7	1.104
мчаться	498.5	168.7	1.104
месяц	637.0	298.3	1.103
плих	304.6	0	1.102
чаша	512.3	187.6	1.102
муха	387.7	76.5	1.101
15. Iambic trimeter with alternating feminine and masculine endings (I3fm = Я3жм; 69,813 tokens)			
милый	1,776.2	968.2	1.204
она	5,471.8	4,105.2	1.192
.	7,878.2	6,164.1	1.187
жить	2,177.2	1,378.1	1.183
лишь	1,962.4	1,227.4	1.174
я	25,081.3	20,962.7	1.172

Word	Focus corpus freq.	Reference corpus freq.	K
любить	2,105.6	1,366.9	1.169
любовь	2,506.7	1,748.0	1.160
хоть	1,188.9	638.2	1.151
ах	1,045.7	536.2	1.144
петь	1,733.2	1,139.3	1.143
сам	1,618.6	1,064.2	1.136
пять	558.6	138.6	1.134
а	6,460.1	5,351.3	1.133
о	5,171.0	4,240.6	1.128
мой	7,162.0	6,027.0	1.126
нет	2,735.9	2,117.2	1.121
весна	1,203.2	761.0	1.118
окно	1,303.5	855.3	1.116
но	6,617.7	5,616.3	1.116
ода	415.4	68.7	1.113
хотеть	1,317.8	887.8	1.111
семьдесят	343.8	11.8	1.110
друг	2,034.0	1,573.8	1.101

16. Iambic hexameter with alternating masculine and feminine endings (I6mf = Я6мж; 65,649 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
твой	4,737.3	3,552.5	1.181
давно	1,127.2	502.6	1.178
…	2,223.9	1,437.7	1.177
взор	1,416.6	756.6	1.176
я	25,042.3	20,965.2	1.170
их	1,675.6	1,001.1	1.169
и	52,795.9	45,000.9	1.162
мой	7,433.5	6,025.4	1.156
день	3,564.4	2,693.2	1.153
ты	12,445.0	10,541.6	1.141
средь	868.3	399.0	1.138
душа	3,214.1	2,473.1	1.135
где	3,579.6	2,800.3	1.134
сонет	487.4	78.4	1.133
весь	5,224.8	4,269.0	1.131
пора	1,279.5	800.4	1.126
но	6,687.1	5,616.3	1.124
толпа	822.6	409.6	1.121
свой	4,051.9	3,319.8	1.116
любить	1,873.6	1,369.2	1.115

Word	Focus corpus freq.	Reference corpus freq.	K
таинствен-ный	517.9	159.5	1.113
волна	1,294.8	869.8	1.110
прекрасный	822.6	456.0	1.106
себя	1,934.5	1,471.8	1.103
здесь	1,599.4	1,171.1	1.103
мечта	1,081.5	704.1	1.102
лишь	1,660.3	1,230.3	1.102

17. Iambic hexameter with various patterns of feminine and masculine endings (I6f,m = Я6ж,м; 64,109 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
свой	4,944.7	3,312.8	1.259
их	1,996.6	998.6	1.250
они	5,178.7	3,728.7	1.215
который	1,606.6	803.6	1.211
…	2,339.8	1,436.9	1.203
ты	13,196.3	10,535.9	1.197
твой	4,835.5	3,551.9	1.196
сей	1,216.7	527.5	1.195
себя	2,324.2	1,468.7	1.191
и	53,814.6	44,994.2	1.184
когда	2,979.3	2,058.3	1.182
бог	1,887.4	1,152.2	1.177
душа	3,400.5	2,471.7	1.170
толпа	982.7	408.4	1.169
!	4,398.8	3,359.0	1.164
где	3,712.4	2,799.4	1.157
взор	1,325.9	757.4	1.151
вдруг	1,341.5	779.4	1.149
пред	1,123.1	594.5	1.147
он	9,515.0	7,934.5	1.145
иль	1,310.3	788.7	1.138
природа	857.9	401.2	1.134
слеза	1,263.5	773.1	1.130
певец	608.3	209.7	1.124
средь	811.1	399.6	1.121
уж	1,435.1	956.9	1.121
его	1,700.2	1,198.2	1.120
раб	577.1	206.0	1.116
здесь	1,653.4	1,170.7	1.116
внимать	670.7	289.2	1.116

Word	Focus corpus freq.	Reference corpus freq.	K
красота	811.1	415.2	1.116
ум	857.9	457.9	1.116
улыбка	748.7	375.6	1.111
с	11,121.7	9,716.8	1.110
тогда	889.1	529.0	1.102
18. Trochaic tetrameter with feminine endings (T4f = Х4ж; 60,204 tokens)			
-	5,298.7	3,583.7	1.260
у	4,285.4	2,959.4	1.223
этот	3,969.8	2,712.9	1.220
«	2,126.1	1,216.4	1.216
солнце	2,043.1	1,160.2	1.212
море	1,710.8	970.8	1.186
туча	1,046.4	444.7	1.175
ти	531.5	10.4	1.173
ка	631.2	107.2	1.169
песня	1,777.3	1,090.4	1.168
белый	1,760.7	1,090.9	1.164
…	2,126.1	1,439.0	1.155
»	1,694.2	1,117.5	1.140
ах	1,029.8	536.9	1.139
роза	1,029.8	544.2	1.137
к	5,547.8	4,561.0	1.131
небо	2,690.9	2,041.3	1.129
словно	980.0	528.5	1.128
слышать	1,029.8	579.5	1.126
птица	946.8	505.8	1.126
рыцарь	448.5	79.0	1.120
полный	1,013.2	585.4	1.119
девица	465.1	99.5	1.118
это	2,242.4	1,695.3	1.117
лес	1,295.6	859.3	1.113
если	1,478.3	1,024.8	1.113
с	11,128.8	9,717.4	1.111
народ	897.0	507.7	1.111
сердце	3,072.9	2,478.1	1.109
посмотреть	465.1	128.0	1.108
тройка	365.4	36.6	1.108
тихо	897.0	519.2	1.107
только	2,192.5	1,688.7	1.107
птичка	415.3	88.3	1.106

Word	Focus corpus freq.	Reference corpus freq.	K
отвечать	614.6	270.0	1.105
выходить	847.1	483.9	1.104
дама	431.9	116.0	1.101
19. Iambic tetrameter with masculine endings (I4m = Я4м; 55,349 tokens)			
мой	8,726.4	6,018.2	1.300
он	11,093.2	7,925.2	1.290
когда	3,450.8	2,056.1	1.276
один	3,432.8	2,166.4	1.245
день	3,866.4	2,692.3	1.206
я	25,872.2	20,964.7	1.205
звезда	1,879.0	1,052.1	1.204
дом	1,644.1	884.4	1.196
тень	1,770.6	992.0	1.195
кто	3,179.8	2,205.7	1.187
тот	3,794.1	2,770.9	1.177
сон	2,294.5	1,511.3	1.174
свет	2,439.1	1,632.2	1.174
жизнь	3,017.2	2,123.2	1.174
свой	4,372.3	3,318.5	1.167
огонь	1,806.7	1,117.9	1.167
дно	831.1	291.0	1.164
жить	2,095.8	1,380.1	1.163
луна	1,156.3	578.4	1.161
путь	1,626.0	1,001.3	1.156
как	12,484.4	10,453.4	1.151
снег	1,282.8	726.5	1.149
идти	2,222.3	1,545.2	1.149
там	2,493.3	1,779.0	1.149
петь	1,752.5	1,140.3	1.148
тишина	1,047.9	529.7	1.147
твой	4,498.7	3,555.7	1.144
пора	1,337.0	800.7	1.141
земной	1,029.8	563.2	1.131
так	3,830.2	3,039.0	1.131
чтоб	1,897.1	1,336.4	1.129
лес	1,355.0	859.2	1.128
звон	704.6	287.0	1.127
ночь	2,692.0	2,066.8	1.123
стих	1,210.5	751.4	1.122
час	1,463.4	980.3	1.121

Word	Focus corpus freq.	Reference corpus freq.	K
Ру	361.3	2.8	1.119
ночной	867.2	457.6	1.118
пустой	795.0	397.6	1.117
грудь	1,318.9	865.7	1.117
конь	867.2	475.0	1.113
последний	1,120.2	702.2	1.113
труд	831.1	451.5	1.110
ты	12,014.7	10,547.1	1.108
душа	3,053.4	2,475.2	1.106
прощать	722.7	373.8	1.103
пока	668.5	329.2	1.102
и	49,901.5	45,031.1	1.101
нет	2,637.8	2,119.0	1.101
уж	1,355.0	958.0	1.100

20. Iambic hexameter with paired masculine and feminine endings (I6mmff = Я6мммжж; 53,793 tokens)

Word	Focus corpus freq.	Reference corpus freq.	K
ты	19,296.2	10,498.2	1.652
то	6,934.0	3,179.5	1.608
сей	2,379.5	520.6	1.528
твой	6,878.2	3,539.8	1.510
себя	3,346.2	1,462.9	1.422
когда	3,848.1	2,053.7	1.355
свой	5,186.5	3,313.2	1.297
во	2,825.6	1,508.9	1.292
который	1,877.6	802.8	1.283
мой	8,402.6	6,020.9	1.264
свет	2,844.2	1,629.6	1.262
тот	4,182.7	2,768.4	1.245
мочь	2,751.3	1,654.5	1.236
коль	910.9	169.0	1.234
зреть	892.3	182.6	1.223
всякий	985.3	269.4	1.219
чтоб	2,268.0	1,334.0	1.215
быть	9,703.9	7,490.9	1.211
иметь	873.7	208.5	1.207
честь	892.3	224.7	1.207
стих	1,524.4	749.4	1.207
труд	1,134.0	449.5	1.198
сам	1,859.0	1,063.7	1.196
она	5,484.0	4,107.8	1.194

Word	Focus corpus freq.	Reference corpus freq.	K
желать	780.8	174.0	1.191
любезный	706.4	118.4	1.189
что	9,852.6	7,817.2	1.188
чрез	687.8	104.7	1.188
уж	1,691.7	955.8	1.186
к	5,948.7	4,559.1	1.184
слава	1,226.9	572.9	1.183
они	4,926.3	3,732.3	1.177
ум	1,059.6	457.1	1.174
день	3,680.8	2,693.7	1.173
я	25,077.6	20,971.0	1.171
имя	966.7	409.1	1.164
кто	3,048.7	2,206.8	1.162
приятный	576.3	85.0	1.159
сие	576.3	91.5	1.157
тогда	1,078.2	528.2	1.156
отрада	632.1	143.9	1.155
век	1,394.2	813.3	1.152
утеха	539.1	72.9	1.152
лишь	1,859.0	1,229.6	1.149
другой	1,468.6	888.6	1.149
хотя	669.2	206.7	1.144
судья	501.9	61.9	1.144
муж	632.1	181.4	1.142
счастливый	836.5	361.3	1.141
час	1,543.0	979.9	1.141
достойный	520.5	85.4	1.141
велеть	539.1	105.4	1.140
любовь	2,416.7	1,750.1	1.140
его	1,784.6	1,198.3	1.140
весь	5,279.5	4,270.0	1.139
уже	1,264.1	751.3	1.137
ее	1,245.5	739.9	1.135
писать	725.0	288.8	1.133
муза	687.8	257.5	1.132
внимать	725.0	289.3	1.132
всегда	910.9	458.4	1.131
хоть	1,115.4	639.8	1.131
российский	446.2	48.0	1.131
<	520.5	113.5	1.131
несчастный	520.5	117.2	1.129

Word	Focus corpus freq.	Reference corpus freq.	K
монархиня	390.4	4.9	1.128
б	1,226.9	746.3	1.128
ль	1,264.1	786.2	1.126
величество	390.4	17.2	1.124
представлять	427.6	55.5	1.122
глас	520.5	139.7	1.121
иль	1,245.5	789.8	1.120
днесь	409.0	42.6	1.120
видеть	1,747.4	1,242.7	1.119
ко	818.0	414.4	1.118
сатира	390.4	31.4	1.118
должный	594.9	215.9	1.118
любить	1,877.6	1,369.9	1.116
надпись	409.0	56.7	1.115
счастие	1,226.9	789.8	1.115
прекрасный	855.1	456.3	1.115
покой	818.0	426.4	1.114
везде	520.5	161.9	1.113
рок	501.9	147.2	1.113
один	2,751.3	2,171.3	1.112

Word	Focus corpus freq.	Reference corpus freq.	K
ныне	539.1	181.6	1.112
являть	409.0	69.7	1.111
толь	353.2	19.6	1.110
он	9,146.2	7,939.0	1.110
взор	1,171.2	759.2	1.110
кой	427.6	88.7	1.110
плод	501.9	154.3	1.110
закон	557.7	208.1	1.109
красота	780.8	415.9	1.107
стараться	390.4	66.7	1.106
мучение	409.0	82.2	1.106
взирать	390.4	68.0	1.105
но	6,525.0	5,619.0	1.105
наука	446.2	120.5	1.104
пиит	334.6	22.8	1.103
град	446.2	128.5	1.102
забава	409.0	93.4	1.102
краса	539.1	214.8	1.101
польза	353.2	45.9	1.101
можно	539.1	216.3	1.100
догадка	316.0	14.1	1.100