

The fall of genres that did not happen: formalising history of the “universal” semantics of Russian iambic tetrameter

Antonina Martynenko, Artjoms Šeļa*

Abstract: This paper examines thematic features of the “universal” poetic meters that are disproportionately popular in a tradition. Focusing on the example of 19th-century Russian iambic tetrameter, we propose a method for diachronic analysis of semantic structure of a meter based on the combination of topic modelling and network analysis. We represent each poem as topic probabilities and use most probable topics appearing within one poem to build up the connections of a network. This representation allows us to detect a chronological process of semantic expansion of the meter: its usage spreads to various thematic domains contributing to its perception as “universal”. At the same time, we show that the expansion of iambic tetrameter does not lead to the collapse of semantic traditions of other meters and their associations with certain genres. Testing the amount of shared connections between meters against randomized data demonstrates that the increase in number of topics within a meter is mostly driven by the sample size, rather than by the direct borrowing from other meters. The stability of thematic connections inside each meter displays the conservative nature of poetic meters and surprisingly strong retention of association between meter and semantics.

Keywords: semantic halo, iambic meters, topic modeling, networks, Russian poetry

Background

In versification studies, the association between meter and meaning, an effect also known as “semantic halo of meter”, is a well-studied area with a long scholarly tradition. First examined qualitatively in rare metrical forms that formed particularly noticeable connection with a set of distinctive themes (Taranovskij 1963; Jakobson 1979 [1938]; Vishnevskij 1985; Gasparov 2012 [1999]; Shapir

* Authors' addresses: Antonina Martynenko, Department of Slavic Studies, Lossi 3, 51003, Tartu, Estonia. E-mail: antonina.martynenko@ut.ee; Artjoms Šeļa, Department of Methodology, Institute of Polish Language, Polish Academy of Sciences, Mickiewicza 31, 31–120, Kraków, Poland. E-mail: artjoms.sela@ijp.pan.pl.

2015), semantic halo effect was also found in a number of European poetic traditions (Talinskaya, Oganosova 1986; Tarlinskaya 1989; Červenka 1991; Gasparov 2012; Dobrzyńska 2014). Further studies demonstrated the effect in large historical corpora in multiple languages via computational modelling (Šeļa et al. 2020, Šeļa et al. 2022).

Previous large-scale studies (Gasparov 2012; Šeļa et al. 2022) show that stable relationship between semantics and form exists not only in less frequent meters, but also in wide-spread forms that became central to a tradition: for example, iambic pentameter in English, trochaic and iambic pentameters in Czech, iambic tetrameter in Russian. Commonly used meters, however, are often described as being “universal”, or semantically “neutral”: potentially any theme can be encountered within them. By formalising and exploring semantics of a meter with computational modelling, we will show that “universality” in metrical usage does not exist as such. It is often conflated with normativity that sets average expectations for what an average poem and average poetic theme should look like.

In this paper we want to address the issue of universal semantic profile and its dynamics in wide-spread forms by focusing on the history of expansion of Russian iambic tetrameter (abbr. as Iamb-4), which is the most “classic” and widespread meter in Russian poetry. Its history in Russian tradition started in the middle of the 18th century with the establishment of the accentual-syllabic metrical system that was transferred from German (Gasparov 2003: 182–186). At the same time, the new way of writing verse was directly associated with the classical genre system where specific meters were expected to be used within specific genre domains¹. The place of the iambic family in this hierarchy was high from the very beginning. Iambic meters were proclaimed to be the most suitable for literary versification in Russian and the use of iambic tetrameter was allocated for the ode – doubtlessly the most important genre in Russian 18th-century literature (Tynianov 1977 [1922]; Shapir 1996). In the end of the 18th century poetry started to change its prescribed setting: from the court and public-oriented rhetoric it moved to a meditative introspection, which changed the usage of iambic forms as well. Iamb-4 started to associate with the emerging genres of elegy and fragment (Gasparov 2000: 58–62). Thus, from holding the highest place in the neoclassical hierarchy Iamb-4 allegedly

¹ Genre and meter association generally was not prescriptive but existed in poetic practice since the 18th century (Gasparov 2000: 57–59), see Pumpjanskij (1937) for ode specifically. In the early 19th century one can find explicit reflections on the connection between Iamb-4 and ode or iambic hexameter and elegy, e.g., in the *Dictionary of Ancient and New Poetry* (Ostolopov 1821, 1: 370–371; 3: 491).

became the form that disrupted that system. Nevertheless, these historical generalisations depend on the few overview studies and are not elaborated; research in versification was traditionally focused on the rhythmical aspect of Iamb-4 and its evolution (see, for example: Taranovskij 2010 [1953]; Shapir 1996; Dobritsyn 2016; Liapin 2016; Liapin 2020; Tver'yanovich 2022).

The history of Iamb-4 was situated in a wider context by Mikhail Gasparov who directly connected the expansion of Iamb-4 with the disruptions in the genre system in the 1820s. According to Gasparov, Russian poets of the 1820s switched from the old genre-associated meters to Iamb-4 making it a genre-neutral and universal form, which led to the collapse of the system of classical poetic genres, and, ultimately, resulted in the decline of Iamb-4 itself (Gasparov 1974: 36–76; Gasparov 2000: 112–114, 166–167). In the 1840s poets rediscovered trochaic and ternary meters and established new associations between forms and themes. Thus, the period of Iamb-4's expansion (around 1820s) separates two major epochs in the history of Russian poetry: the late classic system of strong association between genres and the modern one, where genres played a lesser role in dictating the association between themes and meters.

The history of Iamb-4 suggests a diffusion of semantic halo effect, since the form did not retain connections with particular themes, but, allegedly, had lost them. We argue that this diffusion process is typologically the same meter-meaning association phenomenon. The universalisation of a meter's semantics can be seen as the process of tradition's development from a specific narrow thematic domain to a wider pool of themes, which nevertheless remains dependent on usage of the meter in the past.

In order to examine the historical dynamics of meter-meaning association we infer themes of each poem in a corpus using a topic modeling algorithm and propose an exploratory networks-based method for visualisation and examination of themes connectivity inside a meter. Analysing networks' structures on a timescale, we aim to investigate how meter-meaning relationships are structured in universally used meters. We will detail the chronology and features of iambic tetrameter's expansion towards new themes.

We expect that it is possible to trace a meter's semantic expansion using computational modelling, as the universality of a meter can be understood as its thematic versatility. At the same time, we assume that the association between meter and semantics established during the classical period (before 1820s) will not fade away during the process of Iamb-4 change and diffusion. First, we assume that the expansion of one meter did not disrupt other forms that continued to be used within their thematic domains. Second, we expect that it is possible to test Iamb-4's influence on other meter's semantic halos directly. In particular, we will examine the effect of a sample size as a cause for thematic versatility versus the hypothesis of thematic borrowings between meters.

Although this article takes the development of Russian meters as the main object of investigation, we claim that the theoretical and methodological framework we use is applicable for modelling the history of universal meters in other poetic traditions (see an example from Czech poetry in (Plecháč, Kolár 2022)).

Data

Two corpora are used in the study². Main analysis is performed using the poetic subcorpus of the Russian national corpus (RNC) in a time frame between 1775 and 1850. The median length of a poem in this period is 22 lines, though the corpus includes many longer narrative poems with size about 500–2000 lines (the longest document³ in the corpus is Pushkin’s “*Ruslan and Ljudmila*” which contains 2799 lines). To make text lengths comparable and compatible with topic modeling, poems longer than 25 lines were sampled and presented as fragments of 22 consecutive lines: we take one sample per each of increasing size intervals (25, 50, 100, 200), with a maximum of 6 samples for texts longer than 300 lines. This way we adapt the corpus for topic modeling and also allow longer texts to be represented with more textual evidence. As the initial subset contains 10 552 unique poems, they resulted in 12 830 fragments after sampling. That includes more than 970 000 words and about 35 000 unique words. We refer to this corpus as the “canonical corpus”, or RNC.

The sources of the texts in RNC are 20th-century critical editions, which creates a bias in this collection towards the literary and academic canon of Russian poetry of the 19th century (Korchagin 2015). As a less biased contrast corpus we use a small historical collection of lyrical poems printed in periodicals between 1835 and 1839 (referred as the “periodicals corpus”, or PC). The corpus includes 897 poems (1191 fragments after sampling) and about 116 thousand tokens (~11 thousands types), no author or genre-based selection was performed. We use this set for two reasons: 1) it is a “control” corpus against which we check the representativeness of the canonical collection; 2) we aim to validate Gasparov’s claim that poems appeared in the second half of the 1830s might be of particular importance in terms of Iamb-4 destabilisation and loss of thematic versatility.

² See the code and data published at: <https://doi.org/10.5281/zenodo.7958274>

³ Longer narrative poems indeed exist in this period, though they were separated into multiple documents (by chapters) in the corpus by default.

The direct comparison between two collections is, nevertheless, difficult. Most texts in the RNC have the dates of creation in metadata, as opposed to the periodical collection, where only the publishing year is available. Seeing these and other⁴ issues in direct comparison of the two corpora, we selected a sample of 1 173 fragments dated as 1830s from RNC to be compared to PC.

Texts in both corpora were supplied with metadata: most importantly, the date of a text's creation or publication and its metrical features (meter and feet length). Although we build the topic model on the whole corpus, only four metrical forms reach stable frequencies during the whole period that are sufficient for network analysis⁵. All texts were preprocessed, i.e., transformed to the modern Russian orthography and lemmatised via *pymystem3* (Segalovich 2003).

Methods

Topic modelling

How to infer thematic features in a collection of texts? Working with poetic texts, one needs an approach that goes beyond the frequencies of individual words which might be biased by various factors (morphology, meter, genre) and, generally, are not very useful, or informative, in short texts. We rely on topic modeling to find more abstracted groups of words that share similar behavior, appear in similar contexts with each other, which often is a signal of an underlying shared theme, or semantic domain. Topic models are widely used in information extraction tasks and were multiple times proven to be a robust method of inferring themes in poetry (Navarro-Colorado 2018; Plecháč, Haider 2020), including its application to the problem of semantic halo (Šeĭa et al. 2022; Plecháč, Kolár 2022).

We use Latent Dirichlet Allocation: a popular algorithm based on a generative assumption, which allows us to represent each text in a collection as a mixture

⁴ In addition to date annotation incoherence, the imbalance of the canonical corpus of the 1830s prevents us from the direct comparison of RNC and PC. RNC is overpopulated by the poems by major authors: 45% of poems between 1835 and 1839 are those by Mikhail Lermontov, Aleksandr Pushkin, Vladimir Benediktov, Vasilij Zhukovskij and Aleksandr Polezhaev. The periodicals corpus presents a more even distribution of poems per author with no author presenting more than 3% of the texts in the corpus. Thus, for the comparison we sampled texts from RNC, controlling the number of poems per author up to 30 and taking the texts dated in RNC between 1830 and 1839.

⁵ See the number of texts in each meter in Appendix A2.

of topics, as if each text was “generated” from all available topics with varying probabilities. We trained a model with 75 topics⁶. Each text became represented as a probability distribution over these 75 topics: the model placed all texts in the same thematic space and made them directly comparable. Each modelled topic contains all words from the corpus, though only a small number of words are highly probable. To give an example of two topics’ most probable words:

Topic 49: grob, mogila, prakh, zemlja, mërtyvj [tomb, grave, ash, soil, dead]

Topic 69: tsar’, narod, russkij, tsarstvo, knjaz’ [tsar, nation, russian, tsardom, prince]

One can infer that Topic 49 may be encountered in graveyard elegies, while Topic 69 is more probable in patriotic, historical, or folk poems (or at the intersection of all of these). This example shows that, although the model is built solely on the lexical frequencies and has no knowledge of genre or meter, it succeeds in grouping words in distinct genre-related topics. To examine the topic association with particular meters, we can identify the most probable topics in each text and then aggregate topic probabilities according to texts’ meters. Finally, we aim to analyse the *structure*⁷ of thematic profiles in meters via network representation.

Exploration with networks

Topic-based representation helps to identify sets of the most probable topics appeared in each poem. If two or more topics appear together in one poem with a high probability, we see these topics as *connected*. Our approach is summarized on Fig. 1:

Step 1: selection of the most expressed topics in a poem;

Step 2: drawing connections between these selected topics (i.e., present topics as nodes and draw an edge if the topics are connected) in each poem;

LDA is a generative algorithm and assigns each document with probabilities of all topics, with most of the values being very low, as demonstrated on Fig. 1. To select only those topics that are meaningfully present in a text, we consider only the highest 5% of probabilities from the overall probability distribution.

⁶ For more details about the LDA model see Appendix A1.

⁷ Compare with Mikhail Gasparov’s opinion on semantic halos being not just a set of themes but their grouping in a specific structure: (Gasparov 2012: 159).

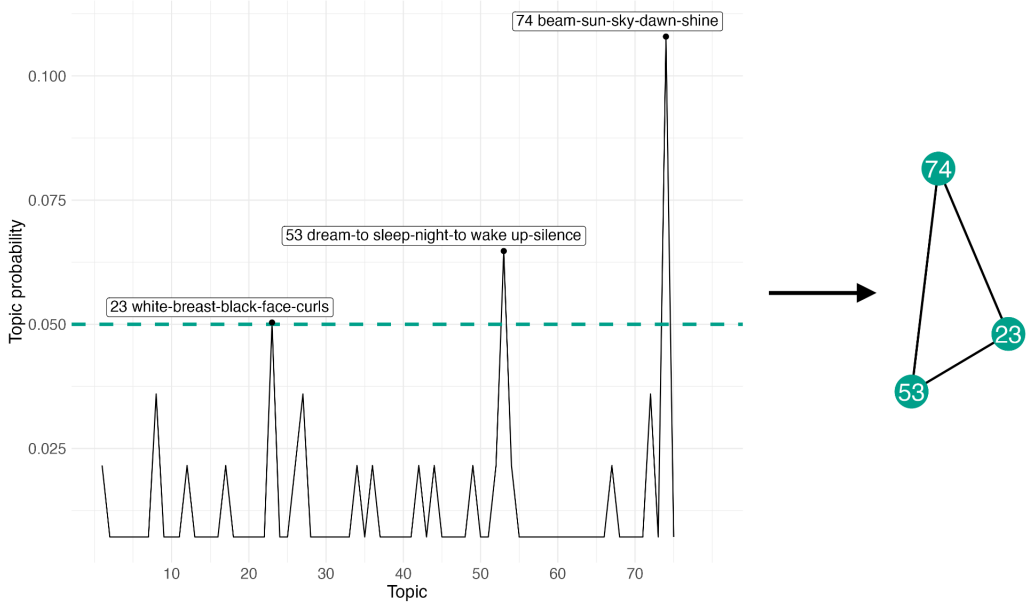


Figure 1. Example of topic selection from the most probable topics of one poem (“Insomnia” (1838) by Nadezhda Teplova) that results in a network with three connections; green dashed line is the threshold.

After we calculate the most probable topics and connect them inside each poem, we aggregate connections by meters and time periods:

Step 3: aggregation by groups, i.e., summarising connections for each meter, for each 5-year time periods;

For instance, the poem shown on Fig. 1 resulted in three connections (23–53, 23–74, 53–74) added to iambic hexameter, time slice between 1835 and 1839.

Step 4: construction of networks, where topics are nodes and edges are topic connections that are weighted according to the data aggregated in Step 3.

To explore the data we suggest using network visualisations. We present each topic as a node and build edges in cases when two or more topics were identified as connected, i.e., highly probable in the same text. However, we filter out edges that appear less than two times in a respective period to cull out noise.

An example of a meter’s semantic network is presented on Fig. 2. To build the network we summarise all the connections for iambic hexameter poems dated

between 1775 and 1849 (for the clarity of demonstration we use only connections that appeared at least 4 times). We used a clustering algorithm (Louvain method) on the resulting network to automatically detect communities. The aim of the clustering is to find the nodes (topics) which appear in a similar context to group them together, or, on the contrary, to separate those that rarely appear with each other. Neither of the algorithms we used was given any information about genres of the poems. Nevertheless, if we read the topic united in clusters on Fig. 2, they appear to be clearly related to traditional genres of iambic hexameter, such as idylls, elegies, and didactic poems (cf. Gasparov 2000: 62–64, 117–118). This result lets us assume that the formalisation and modelling steps we performed indeed provide a reasonable and human-readable output that is comparable with qualitative assessments of meters' semantics.

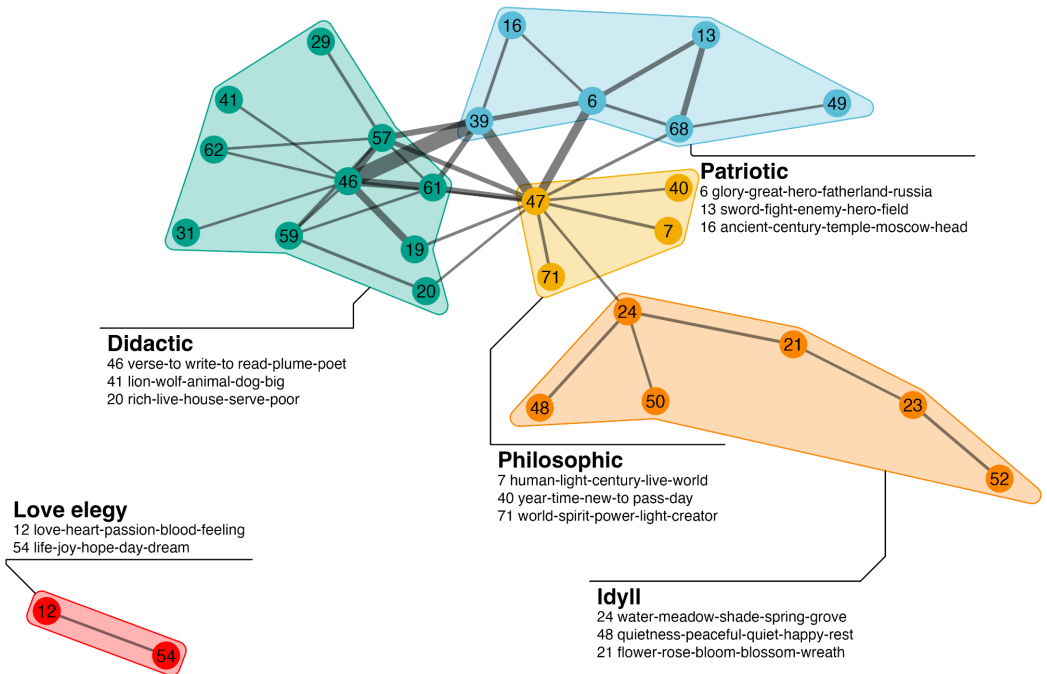


Figure 2. Resulting network example: most frequent connections for iambic hexameter (1775–1849) grouped by a community detection algorithm (Louvain method); we labelled communities manually, below each label we provide selected topics with five most probable words in each; width of the edge set as the connection's frequency.

Results

Network expansion

Thematic networks of each meter demonstrate a significant difference between Iamb-4 and all other meters' historical dynamics (see Appendix A4). While most meters typically occupy only a particular semantic domain (i.e., establish connections only in a specific part of the network), Iamb-4 clearly expands to the whole network space, even if for a short period (Fig. 3).

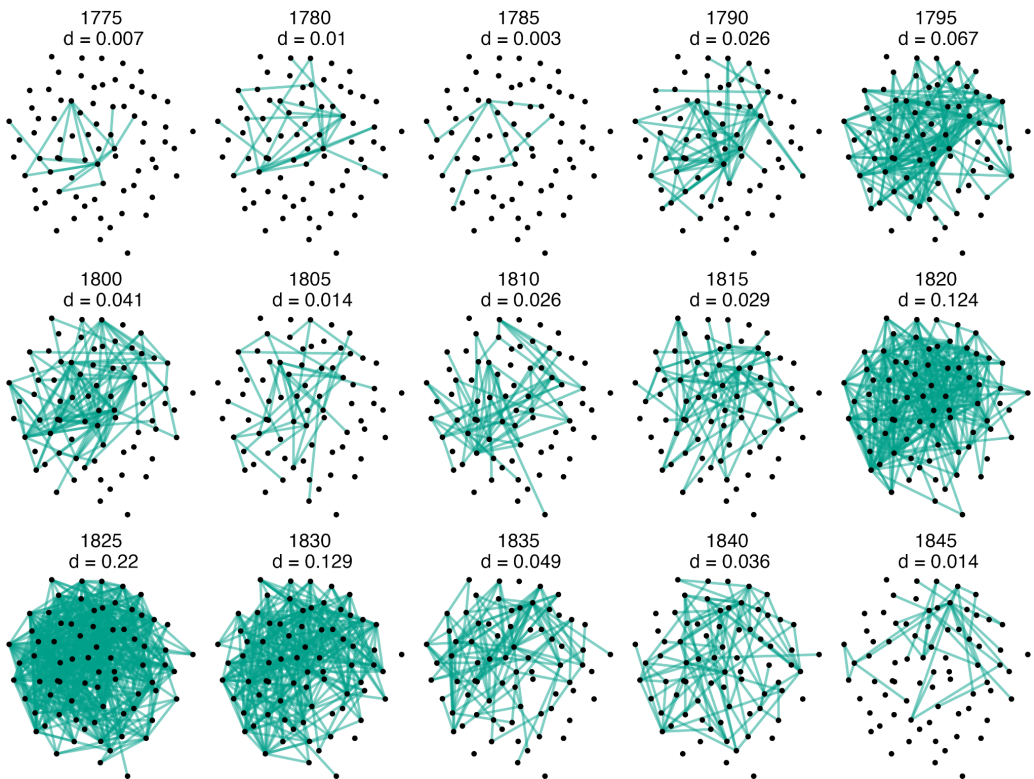


Figure 3. Networks of Iamb-4 in each 5-year time slice; d is the network density coefficient.

For instance, from having only 81 connections between 1815 and 1819 Iamb-4 expanded to 343 in the next 5-year slice; then, between 1825 and 1829, it reached the maximum number of 610 edges. In relative values, the network density scores (number of present connections to the ratio of all possible connections) of the Iamb-4's networks between 1820 and 1835 are dramatically

higher than any network of any meter (including Iamb-4) ever achieved in this period⁸.

The process of Iamb-4 expansion suggests that some topic connections emerging in Iamb-4 in the 1810s and 1820s might previously exist in other meters and then were transferred to Iamb-4. Iambic hexameter can serve here as an example. In the beginning of the 19th century this meter was mostly used either for didactic genres or for idylls and elegies (Fig. 2). Among “didactic” themes the meter has one very frequent connection: a meta-poetic theme related to writing, inspiration, and the nature of poetry (edge between topics 19 and 46). The same connection became frequent in Iamb-4 in the period between 1815 and 1819, i.e., right before the thematic expansion visible from the networks (see Tab. 1). Similarly, some of frequent meta-poetic connections present in the early 1820s in iambic verse with irregular feet length (the “free” iamb)⁹ can be found in the Iamb-4’s network in the 1820s.

Thus, it seems that the thematic “expansion” of Iamb-4 follows the use of this meter for reflection on writing poetry: a theme, previously associated with other iambic forms. Meta-poetic topics also appear in the most common poetic forms in modern European poetry in different languages (see Šeļa et al. 2022, Table S3), which makes these meters emblematic of a tradition as a whole.

Table 1. Examples of shared frequent (among top 5) connections between Iamb-4 and other meters in respective periods of time

Connection presence		Edge ID	Most probable topic words	
Iamb-6 1805-1819	Iamb-4 1815-1819	19-46	19: muza, pevets, slava, dar, poet [muse, bard, fame, talent, poet]	46: stikh, pisat', chitat', pero, poet [poem, to write, to read, plume, poet]
“Free” iamb 1820-1824	Iamb-4 1820-1829	19-42	19: muza, pevets, slava, dar, poet [muse, bard, fame, talent, poet]	42: dusha, mechta, poet, zhivoj, vostorg [soul, dream, poet, vivid, delight]
Iamb 4 1825-1829	Trochee-4 1835-1839	26-66	26: volna, more, bereg, voda, breg [wave, sea, shore, water, strand]	66: burja, tucha, grom, nebo, veter [storm, thundercloud, thunder, sky, wind]

⁸ According to our calculations, the mean network density for the built networks is about 0.02.

⁹ Iambic verse with irregular line length (number of feet per line) that is perceived in Russian tradition as a separate meter (see Matjash 2011).

The trochaic tetrameter is a reversed example: the meter became more popular after the peak of Iamb-4's popularity (cf. Appendix A2) and might have acquired some of the Iamb-4 frequent connections. Among top 5 most frequent connections in each period, there is at least one edge, which is present both in Iamb-4 and trochaic tetrameter: clearly Romantic, sea and storm related connection (26–66) that both described the space of the sublime and served as a metaphor for abundance of emotions. At first, it is visible in Iamb-4 in the second half of the 1820s, in texts like Pushkin's "The Storm"¹⁰ and those that followed Pushkin's Romantic poems (e.g. "The Corsair" by young Lermontov (1829)). Later the edge 26–66 appears as one of the top 5 frequent connections in trochaic tetrameter poems dated between 1835 and 1839: e.g., "The Sea" (1839) by the most popular poet of this period Vladimir Benediktov, or "The Sea Seine" (1839)¹¹ by another prominent poet, Nikolaj Jazykov.

The emergence of sea-related topics was previously detected in a number of Romantic poetic traditions, including Russian (Plecháč, Haider 2020). According to our data, not only an individual sea-related topic has a rising tendency, but a stable connection of two topics ("Sea" and "Storm") is detected: first, in the major meter (Iamb-4) and then in the emerging one (trochaic tetrameter). The latter can be considered a "transfer" of new Romantic themes acquired by Iamb-4 to other meters in the late 1830s and imply the need of a closer look into this period's data.

1830s: Canon vs. Archive

According to chronological networks of the canonical corpus, the second half of the 1830s was indeed the time when Iamb-4 lost a significant part of this thematic connections (Fig. 3), but this might be just an outcome of poor RNC representation of this meter. Going beyond the canon, we can directly compare whether the same features can be found in the networks built on the less biased data that was collected from the periodicals (Fig. 4).

¹⁰ Cf.: "Hast thou seen on the rock the maid, / In robe of white above the waves, / When seething in the storm dark / Played the sea with its shores" (Pushkin 1888: 109) ["Burja" ("Ty videl devu na skale...") (1825)].

¹¹ Titles of poems in Russian: V. Benediktov – "More" ("Vikhor! Vzryv!—Gigant prosnulsja..."); N. Jazykov – "Morskaja tonja" ("More jasno, more bleshchet...").

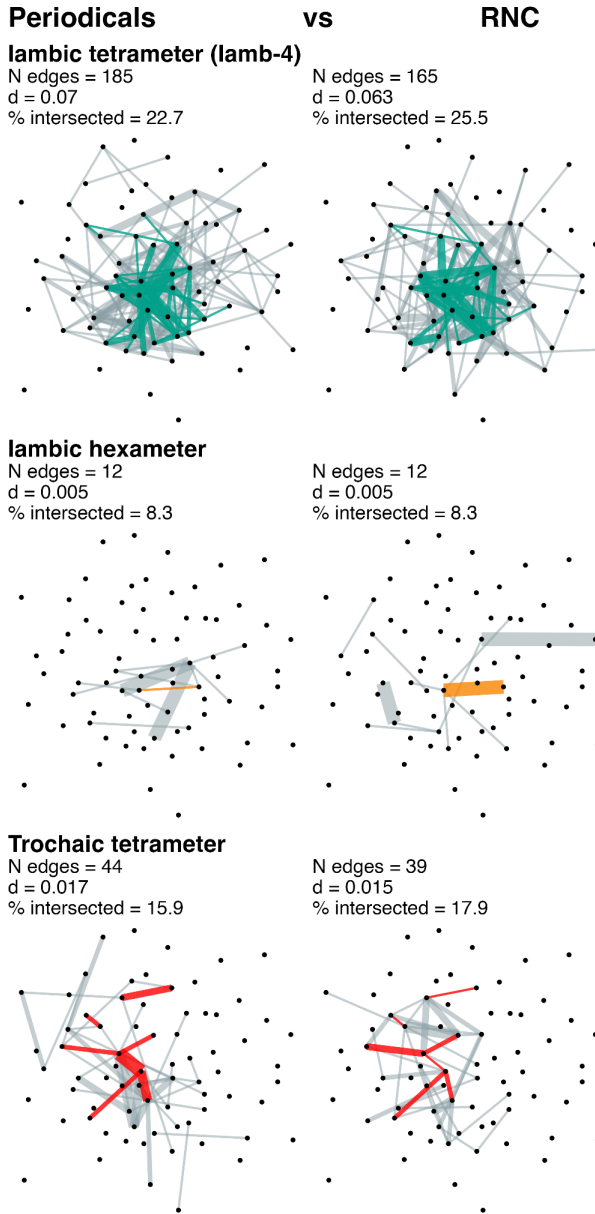


Figure 4. Networks built on the sample from canonical corpus (right) of the 1830s and periodicals corpus (left). Coloured (non-gray) edges are those shared between the two corpora, the percentage of shared edges out of total (N edges) shown as “% intersection”; d stays for network density; width of the edge reflects the connection frequency.

In the case of Iamb-4, the network built on the periodicals data is consistent with the sample from the canonical corpus, having roughly the same number of edges and network density. This result corroborates Iamb-4's thematic variability that couldn't be simply a result of the selection of texts in RNC.

At the same time, the periodicals data for trochaic tetrameter displays more consolidated and stable connections in comparison with the sample from the canonical corpus. In particular, periodicals network reveals three nodes with high degree and betweenness centrality uniting nature, water, and rural topics (74, 26, 67¹²) which only partially overlaps with nodes in the canonical network (most central nodes: 13, 1, 74, 26).

Similar difference is visible in iambic hexameter (see Fig. 4). Although the number of poems in this meter is low in both corpora, the data obtained give evidence for canonical selection leading to a slightly less centralised network with isolated connections. This suggests that the representation of less popular meters such as trochaic tetrameter and iambic hexameter might be influenced by the canonical version of poetic history, where only a few most distinct exemplary texts in these meters were selected.

Iamb-4 shared connections and random baseline

The most obvious obstacle to the analysis presented above is the highly unequal popularity of meters. Poems written in Iamb-4 heavily outnumber poems written in any other form, both in canonical and periodical corpora (cf. Appendix A2).

On the one hand, this validates the usage of RNC as a representative corpus, since the unbiased sample of printed poetry shows a similar distribution of metrical forms. In the case of Iamb-4, there is also a visible similarity between the RNC and periodicals networks, since they have a comparable number and structure of connections.

On the other hand, both corpora used for the network-based exploration have Iamb-4 as the only dominant metrical form, always present in high frequency. Fundamentally, it remains unclear whether the expansion of Iamb-4 over the network is "thematic" in nature (e.g., themes are reallocated from other meters to Iamb-4), or is just a function of its popularity and frequency.

To test this, we calculate the random baseline for the number of shared connections between two random samples of poems. Each sample is taken

¹² See the full list of topics with their most probable words in Appendix A3.

in a size corresponding to each of the compared meter's frequency in a given time period. In other words, we assume that with an increase in sample size, the number of shared connections also naturally grows by sheer chance. If there is, indeed, an expansion and reformation of traditional meter semantics, a number of shared connections that we actually see in the data should be above the random baseline.

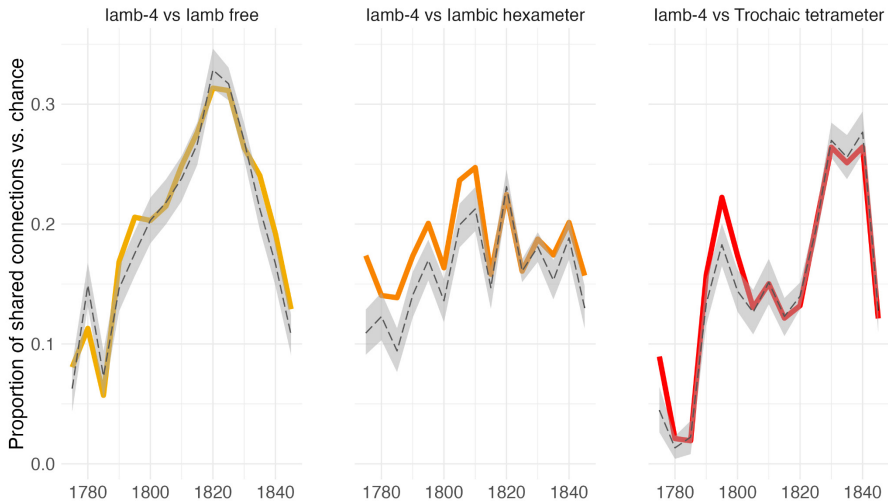


Figure 5. Number of shared connections between Iamb-4 and other meters (in colour) in comparison with random baseline (in grey) given the sample size. Shaded lines correspond to a bootstrapped confidence interval (1000 iterations per given pair of sample sizes).

As Fig. 5 demonstrates, in most cases the number of shared connections can be interpreted purely as a function of Iamb-4 frequency, since the true number of shared connections (in colour) barely extends outside the space of the random baseline (in grey).

Moderately higher number of shared connections can be seen between Iamb-4 and iambic hexameter before 1810, giving more validity to the reallocation of some connections between these forms described above. In general, we see that all slight deviations observed on Fig. 5 happen in the earlier period of the Iamb-4's history (before 1810), when all of the four meters under consideration were used in their respective genre domains. Although the scholarly tradition suggests Iamb-4 to expand to these genres and collapse them, we argue that it did not happen. By and large the expansion of Iamb-4 can be interpreted only as a function of the number of poems written in this meter.

The latter is supported by the analysis of stability of most prominent edges across meters. If thematic expansion of meters leads to the loss of semantic halo, we should also see the loss of meter’s frequent connections over time. At the same time, we would expect Iamb-4 to accumulate connections towards the 1820s that do overlap with other meters’ frequent connections.

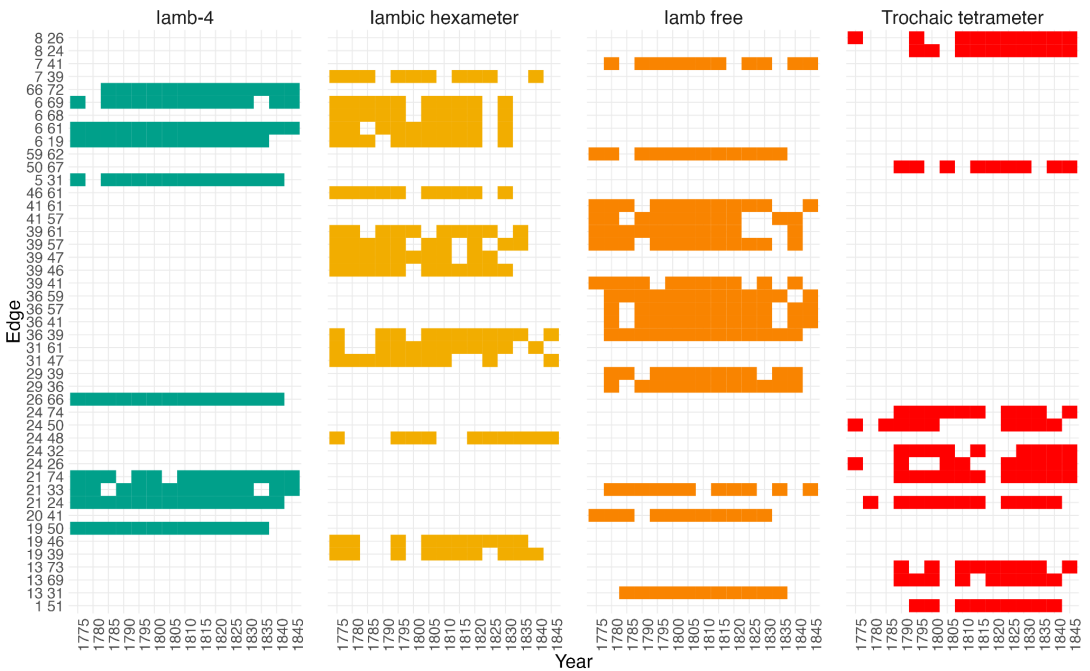


Figure 6. Top 10 most frequent connections in each meter placed on the timeline (in case of a tie, ties are also displayed)

We see exactly the opposite when the most frequent connections in each meter are plotted across time (Fig. 6). Despite the hypothesis that Iamb-4 borrowed thematic connections from other meters roughly between 1810 and 1830, it seems more plausible that these are Iamb-4’s own connections that were established in the end of the 18th century that remained stable during the whole period. Roughly similar dynamics is found in other forms, e.g., in trochaic tetrameter: in most cases its most frequent topic connections are stable between 1800 and 1850.

Yet, the connections for iambic hexameter and free iambic verse show a slightly different dynamic. More robust connections in these meters are found in the beginning of the period, while some topic connections faded between 1820 and 1840. This process is especially visible in iambic hexameter that

shares the same most frequent connections with iambic tetrameter (see, e.g., edges 6–61, 6–69, and 6–19), giving some evidence of these themes reallocation to Iamb-4. It should be said that this example is an exception from the overall distribution of most frequent connections inside meters which are, in most cases, unique and stable despite the small size of some samples.

Discussion

The network-based analysis validates the previous vision of Iamb-4's expansion into different thematic zones in the 1820s and 1830s. As the network densities in this period significantly exceed the density of any network of any other meter, we were able to find formal evidence of the emergence of the universal meter. We see that this result is not a consequence of canonicity bias in the analysed corpus, as testing a comparable set of poetic texts collected from periodicals without selection presents a roughly similar picture. However, we see that there might be a bias in historical representation of semantics of less frequent meters, where filters of literary and academic canon become more pronounced.

Nonetheless, the expansion of Iamb-4 is closely associated with its frequency. The number of connections it shares with other meters generally does not exceed the number achievable by chance and depends on the rise and fall of the meter's popularity. On average, Iamb-4's thematic versatility can be primarily explained by the large number of poems written in this meter, and not by direct invasions into semantic domains of other meters: at least we don't see it systematically with the current methodology. The causes of dramatic increase in popularity in Iamb-4, or other dominant forms across European traditions, are beyond the scope of the paper and should be addressed separately. Our observations indirectly suggest that functional features of a meter, like its "fit" to a language, is not enough to explain its popularity: additional forces, like fashion dynamics (Klimek et al. 2019), elite and generational turnaround (Underwood et al. 2022) might be responsible for the pattern.

The analysis of the stability in topic connections suggests that changes in meters' semantics were not rapid as the most frequent connections are very persistent. Consequently, semantic structure, that is embedded early in a meter, does not disappear, when a meter rapidly expands, like in the case of Iamb-4. This highlights strong inertia and historical continuity in poetic practices at a large-scale: a push that genre conventions give to perceived meaning of meters holds for a long time. We argue that the "universality" of a meter is also a part of its distinct semantic halo, with its thematic profile strongly associated with meta-poetic themes.

These conclusions should be treated with a certain level of caution, caused by the systematic effects meters have on language and lexis (Gasparov, Tarlinskaja 2008). As the LDA algorithm exploits word co-occurrences inside texts written in different meters, there is no guarantee it does not return back topics biased by prosodic affordances within these meters. For instance, Fig. 6 displays little overlap in connections between trochaic and iambic meters: is it a strictly cultural, or strictly linguistic effect? The answer is, most likely, highly unsatisfying: it is both. Meters that are more *linguistically* different also can be more *culturally* different and acquire distinct histories of usage, because there is more chance that humans and poets remember, process and learn them differently. Isolating language and culture in meters is challenging, since the influence of meter is pervasive in all aspects of natural language, from prosody to syntax and semantics, but it must be done in future to get more satisfying answers. What we see now in poetic meters is unwavering continuity and strong dependency of a tradition on its past states.¹³

References

- Červenka, Miroslav 1991. *Z večerní školy versologie sémantika a funkce veršových útvarů*. Pardubice: Akcent.
- Dobritsyn, Andrei 2016. Rhythmic entropy as a measure of rhythmic diversity (The example of the Russian iambic tetrameter). In: *Studia Metrica et Poetica* 3 (1), 33–52. <https://doi.org/10.12697/smp.2016.3.1.02>
- Dobrzyńska, Teresa 2014. Verse forms as bearers of semantic values. In: *Studia Metrica et Poetica* 1 (2), 103–128. <https://doi.org/10.12697/smp.2014.1.2.05>
- Gasparov, Mikhail L. 1974. *Sovremennyj russkij stikh: metrika i ritmika*. Moskva: Nauka.
- Gasparov, Mikhail L. 2000. *Oчерk istorii russkogo stikha*. 2nd ed. Moskva: Fortuna Limited.
- Gasparov, Mikhail L. 2003. *Oчерk istorii evropejskogo stikha*. 2nd ed. Moskva: Fortuna Limited.
- Gasparov, Mikhail L. 2012 [1999]. *Metr i smysl: Ob odnom iz mekhanizmov kul'turnoj pamjati*. Moskva: Fortuna EL.

¹³ The study was supported by the “Large-Scale Text Analysis and Methodological Foundations of Computational Stylistics” project (SONATA-BIS 2017/26/E/HS2/01019).

- Gasparov, Mikhail L.; Tarlinskaja, Marina 2008. The linguistics of verse. In: *The Slavic and East European Journal* 52 (2), 198–207.
- Grün, Bettina; Hornik, Kurt 2011. Topicmodels: An R package for fitting topic models. In: *Journal of Statistical Software* 40 (1), 1–30. <https://doi.org/10.18637/jss.v040.i13>.
- Jakobson, Roman 1979 [1938]. Toward a description of Mácha's Verse. In: Jakobson, Roman. *Selected Writings. Vol. 5*. The Hague, Paris, New York: Mouton, 433–485.
- Klimek, Peter, et al. 2019. Fashion and art cycles are driven by counter-dominance signals of elite competition: Quantitative evidence from music styles. In: *Journal of The Royal Society Interface* 16 (151), 31.07.2018. <https://doi.org/10.1098/rsif.2018.0731>
- Korchagin, Kirill M. 2015. Poezija XX veka v Poeticheskom podkorpuse Natsional'nogo Korpusa Russkogo Jazyka: Problema reprezentativnosti. In: *Trudy Instituta russkogo jazyka im. V.V. Vinogradova* 6, 235–257.
- Liapin, Sergei 2016. The Russian iambic tetrameter: The problem of description (Prolegomena to a new paradigm). In: *Studia Metrica et Poetica* 3 (2), 99–127. <https://doi.org/10.12697/smp.2016.3.2.04>
- Liapin, Sergei 2020. Russian iambic tetrameter: The evolution of its rhythmic structure. In: *Studia Metrica et Poetica* 7(2), 7–22. <https://doi.org/10.12697/smp.2020.7.2.01>
- Matjash, Svetlana 2011. *Vol'nyj jamb russkoj poezii XVIII–XIX vv. Zhanr, stil', stikh*. Sankt-Petersburg: Filologičeskij fakul'tet SPbGU.
- Navarro-Colorado, Borja 2018. On poetic topic modeling: Extracting themes and motifs from a Corpus of Spanish Poetry. In: *Frontiers in Digital Humanities* 5, 15. <https://doi.org/10.3389/fdigh.2018.00015>
- Ostolopov, Nikolaj 1821. *Slovar' drevnej i novoj poezii*. Saint Peterburg: Imperatorskoj akademii nauk.
- Piperski, Alexander 2017. Semantic halo of a meter: A Keyword-based approach. In: *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2017"*. Vol. 2. Moskva: RGGU, 342–354.
- Plecháč, Petr; Kolár, Robert 2022. Metre and semantics in the poetry of Czech post-symbolists accessed via LDA topic modelling. In: *Studia Metrica et Poetica* 9 (1), 7–19. <https://doi.org/10.12697/smp.2022.9.1.01>
- Plecháč, Petr; Haider, Thomas N. 2020. Mapping topic evolution across poetic traditions. *arXiv* 2006.15732 [preprint]. <http://arxiv.org/abs/2006.15732>
- Pumpjanskij, Lev V. 1937. Lomonosov i nemetskaja shkola razuma. In: *Zapadnyj sbornik*. Moskva, Leningrad: Izdatel'stvo Akademii nauk SSSR, 157–186.

- Segalovich, Ilya 2003. A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine. In: *Proceedings of the International Conference on Machine Learning; Models, Technologies and Applications. MLMTA'03, June 23–26, 2003*. Las Vegas, Nevada, USA: CSREA Press, 273–280.
- Šeļa, Artjoms, et al. 2020. Weak genres: Modeling association between poetic meter and meaning in Russian poetry. In: *CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020*. CEUR-WS 2723, 13–21.
- Šeļa, Artjoms, et al. 2022. Semantics of European poetry is shaped by conservative forces: The Relationship between poetic meter and meaning in accentual-syllabic verse. In: *PLOS ONE* 17(4), e0266556. <https://doi.org/10.1371/journal.pone.0266556>
- Shapir, Maksim I. 1996. U istokov ruskogo chetyrekhstopnogo jamba: Genezis i evoljutsija ritma (K sociolingvističeskoj kharakteristike stikha rannego Lomonosova). In: *Philologica* 3 (5/7), 69–101.
- Shapir, Maksim I. 2015 [1991]. “Semantičeskij oreol metra”: Termin i ponjatie. In: *Universum versus: Jazyk – stikh – smysl v russkoj poezii XVIII–XX vekov. Tom 2*. Moskva: Jazyki slavjanskoj kul'tury, 395–404.
- Taranovskij, Kirill 1963. O vzaimootnošenii stikhotvornogo ritma i tematiki. In: *American Contributions to the Fifth International Congress of Slavists*. The Hague: Mouton, 287–332.
- Taranovskij, Kirill 2010 [1953]. Russkie dvuslozhnye razmery. Son'kin, Viktor (Trans.). In: *Russkie dvuslozhnye razmery. Stat'i o stikhe*. Moskva: Jazyki slavjanskoj kul'tury, 11–364.
- Tarlinskaja, Marina; Oganosova, Naira 1986. Meter and meaning: The semantic “halo” of verse form in English Romantic lyrical poems (iambic and trochaic tetrameter). In: *The American Journal of Semiotics* 4 (3), 85–106. <https://doi.org/10.5840/ajs198643/422>
- Tarlinskaja, Marina 1989. Meter and meaning: Semantic associations of the English “Dolnik” verse form. In: *Style* 23 (2), 238–260.
- Tver'yanovich, Kseniya 2022. On the expected and actual rhythmical grammar of Russian iambic tetrameter. In: *Tackling the Toolkit: Plotting Poetry through Computational Literary Studies*. Prague: Institute of Czech Literature of the Czech Academy of Sciences, 179–192. <https://doi.org/10.51305/ICL.CZ.9788076580336>
- Tynjanov, Jurij N. 1977 [1922]. Oda kak oratorskij zhanr. In: Tynjanov, Jurij N. *Poetika. Istorija literatury*. Kino, 227–252.
- Underwood, Ted, et al. 2022. Cohort succession explains most change in literary culture. In: *Sociological Science* 9, 184–205. <https://doi.org/10.15195/v9.a8>
- Vishnevskij, Kirill D. 1985. Ekspressivnyj oreol pjatistopnogo khoreja. In: *Russkoe stikhoslozhenie: Traditsii i problemy razvitija*. Moskva: Nauka, 94–113.

Appendix

A1. LDA implementation details

We used LDA implementation in the *topicmodels* package for R language (Grün, Hornik 2011), training the model of 75 topics with Gibbs' sampling and alpha-prior, controlling the “spikiness” of topic probabilities in a document, set to 0.5. To reduce sparseness of word frequency matrix and boost thematic signal, a list of “functional” stop words was excluded and only 5 thousand most frequent words were used for building the model. All texts in the corpus were used to train the model, so that each poem is presented as a vector of probabilities of 75 topics; each topic itself is a probability distribution over all 5000 MFW in the corpus.

A2. Distribution of texts in the corpus according to metrical forms (after sampling).

Meter / 5-year period	1775- 1779	1780- 1784	1785- 1789	1790- 1794	1795- 1799	1800- 1804	1805- 1809	1810- 1814	1815- 1819
Iambic tetrameter	67	76	65	165	282	207	154	177	228
Iambic hexameter	103	95	53	75	101	70	143	152	108
“Free” iamb	32	182	40	82	104	137	211	236	327
Trochaic tetrameter	22	10	13	85	118	76	63	83	57
Other forms (altogether)	29	20	33	89	97	112	115	239	339
Total	253	383	204	496	702	602	686	887	1059

Meter / 5-year period	1820- 1824	1825- 1829	1830- 1834	1835- 1839	1840- 1844	1845- 1849	Total RNC (1775-1849)	RNC 1830s sample	Periodicals (PC) 1835-1839
Iambic tetrameter	486	733	494	597	308	180	3828 29.8%	329	325
Iambic hexameter	121	119	92	135	129	66	1486 11.6%	82	65
“Free” iamb	183	192	139	407	126	54	2112 16.5%	140	293
Trochaic tetrameter	72	118	165	339	321	102	1462 11.4%	146	156
Other forms (altogether)	322	440	544	394	547	622	3942 30.7%	476	352
Total	1184	1602	1434	1033	1281	1024	12830 100%	1173	1191

A3. List of topics with 5 most probable topic words

- | | |
|--|--|
| 1 starik devica krasnyj molodoj staryj | 38 otec mat' syn rebenok ditja |
| 2 brat dva zhena muzh drugoj | 39 um trud svet nauka glupec |
| 3 iskat' ischezat' vezde dukh sled | 40 god vremja novyj prokhodit' den' |
| 4 deva prekrasnyj krasa chistyj molodoj | 41 lev volk zver' sobaka bol'shoj |
| 5 bog davat' sila bozhe gospod' | 42 dusha mechta poet zhivoj vostorg |
| 6 slava velikij geroy otechestvo rossija | 43 dusha chuvstvo mysl' serdce nemoj |
| 7 chelovek svet vek zhit' mir | 44 zhizn' smert' zhit' umirat' blazhennyj |
| 8 zvezda nebo luna noch' luch | 45 dom skazat' syn bog boginja |
| 9 chas zhdat' prijti poslednij speshit' | 46 stikh pisat' chitat' pero poet |
| 10 pora den' pamjat' pomnit' zabyvat' | 47 kol' sie zret' ljubeznyj smertnyj |
| 11 svjatoj molitva vera khram bozhij | 48 tishina mirnyj tikhij schastlivyj pokoj |
| 12 ljubov' serdce strast' krov' chuvstvo | 49 grob mogila prakh zemlja mertvyj |
| 13 mech boj vrag geroy pole | 50 pet' pesnja zvuk pesn' pevets |
| 14 videt' glaz smotret' edva slyshat' | 51 sidet' okno dver' dom stojat' |
| 15 pit' vino chasha pir gost' | 52 grud' usta sladkij nega sleza |
| 16 drevnij vek khram Moskva glava | 53 son spat' noch' prosypat'sja tishina |
| 17 ljubit' serdce zhit' ljubov' milyj | 54 zhizn' radost' nadezhda den' mechta |
| 18 svoboda volja cep' davat' sila | 55 dusha zhizn' strast' stradanie muka |
| 19 muza pevets slava dar poet | 56 ruka noga golova glaz podymat' |
| 20 bogatyj zhit' dom sluzhit' bednyj | 57 drugoj khotet' inoj stanovit'sja govorit' |
| 21 cvetok roza cvet cvesti venok | 58 den' noch' ten' chas utro |
| 22 slyshat' golos zvuk shum tikho | 59 skazat' stanovit'sja khorosho govorit' |
| 23 belyj grud' chernyj lico kudri | prikhodit' |
| 24 voda lug ten' ruchej roshha | 60 put' idti pojti bezhat' vesti |
| 25 vzor oko rech' vzgljad ulybka | 61 pravda sud zakon delo zlo |
| 26 volna more bereg voda breg | 62 znat' khotet' nichto skazat' bojat'sja |
| 27 serdce дума dusha toska polnyj | 63 drug milyj družhba vernyj schastlivyj |
| 28 schastie sud'ba puskaj nakhodit' | 64 milyj akh zabyvat' proshhat' drug |
| 29 slovo skazat' govorit' znat' nikto | 65 kraj strana rodnoj chuzhoj dalekij |
| 30 nebo zemlja mir zemnoj nebesnyj | 66 burja tucha grom nebo veter |
| 31 vrag smert' strashnyj uzhasnyj krov' | 67 letet' krylo orel ptica letat' |
| 32 vesna pole sneg veter zima | 68 zret' glava glas vnimat' veshhat' |
| 33 krasota prelestnyj milyj prekrasnyj | 69 car' narod russkij carstvo knjaz' |
| nezhnyj | 70 davat' pervyj ruka drugoj chej |
| 34 sleza serdce plakat' toska pechal' | 71 mir dukh sila svet tvorec |
| 35 tolpa svet smekh igra zabava | 72 gora les skala dikij stojat' |
| 36 znat' govorit' točno dva ochen' | 73 kon' doroga pole bezhat' letet' |
| 37 ogon' plamen' grud' goreť oko | 74 luch solnce nebo zarja blistat' |
| | 75 serdce dusha ljubov' dar radost' |

A4. Networks exploration: four meters development on a timeline

