

Computer Models of Verse Prosody¹

Evgeny Kazartsev^[0000-0002-8795-0302]

National Research University Higher School of Economics,
123 Griboyedova Canal Emb., St Petersburg 190068, Russia
kazar@list.ru

Abstract. A new computer system for analyzing verse and prose prosody produces the different probability models of versification used in the hypothetical reconstruction of the mechanisms for generating the verse line. The computer models are created by using natural accents in the language based on the prose rhythm and the specific technique of versification. A correspondence or lack of correspondence between verse and the models provides information regarding the mechanism of versification and the language (prosaic) resources for the poetic rhythm. This research is devoted to the study of verse prosody in different languages in comparison to the models produced by this new computer program for verse and prose rhythm studies. The program is used here to analyze the early iambs in German, Russian and Ukrainian poetry, as well as to observe some differences and similarities among the versification mechanisms in different languages.

Keywords: Metric, Computer Models, Comparative Versification.

1 Introduction

This study is based on comparisons of verse prosody using the different probability models of meter and computer methods for the reconstruction and analysis of text prosody. The focus is on the development of iambic verse in Western and Northern Europe. The addition of Russia to this zone after the Northern war and the reforms of Peter the Great predetermined the fate of Russian and Slavonic literature.² Verse rhythm from the 17th to the 19th century will be studied, with particular attention to the early continental iambs by German, Russian, and Ukrainian poets: Andreas Gryphius, Mikhail Lomonosov, Alexander Sumarokov, and Taras Shevchenko.

This comparative research is based on the analytic system for the study of verse structures and Marina Krasnoperova's theory of reconstructive simulation of versification (RS-theory), in which cognitive and probability models of versification are used

¹ This study is supported by Russian Science Foundation, project no 16-18-10250.

² The problem of the emergence of Russian iambic verse under the conditions of foreign influence has been a central concern in Russian verse studies: specialists were inclined toward the opinion that, despite borrowing its metric structure from German verse, the rhythm of the Russian iamb was formed on the basis of Russian language prosody and did not copy German verse [1].

for the hypothetical reconstruction of the inner processes and mechanisms for generating and perceiving the verse line as they are expressed through its prosodic structure [2, 3]. The RS-theory, which arose at the turn of the 20th and 21st centuries, has been little applied until now, despite the high regard of experts.³

2 A New Computation System and Related Works

In this study, the apparatus of RS-theory takes on a computer embodiment using the platform of a new computing system for analyzing prosodic parameters of texts (CSAPPT), elaborated during the project “Contemporary models of poetics: a reconstructive approach” supported by the Russian Science Foundation. This new system is designed to study the deep processes that accompany the formation of metrical speech prosody. The CSAPPT involves the creation of models for verse rhythm generation that are equivalent to analogous brain processes. We are not aware of other researchers who currently study the problem of text generation based on the methods of metrics and prosody. The modern machine synthesis of speech does not include an adequate cognitive system for modelling rhythmical structure. The development of such models is possible thanks to the analysis of metrically organized texts for which syllable structure has particular significance. There are reasons to think that, at the preverbal level of speech generation, rhythmic “blanks” of commensurate speech segments are formed, which are then progressively filled with words. In metrically organized texts, a preliminary rhythm constitutes a stimulus for all verbal processes. The exploration of this problem takes on a particular relevance considering the challenges currently faced by verse studies.

Currently, the group headed by Igor’ Pil’shchikov is developing a computer program for the metrical analysis of poetic texts; Pil’shchikov is also studying the accentuation of rhythmic words in metrical context [4]. In addition, the National Corpus of the Russian Language includes an expanding poetic subcorpus. Scholars who employ methods for the computer analysis of verse include Boris Orekhov, Sergei Liapin, Robert Kolár and others. The research group of Tatiana Skulacheva and Natalia Sliussar’ is undertaking cognitive studies in the domain of orally performed verse. Skulacheva is also actively studying the linguistics of verse and prose, above all, syntax [5]. Alexandr Prokhorov continues to apply a probability model of verse meter in his studies. Taking into account these studies and initial findings, our research aims at creating an original system, a single platform for various methods and practices of verse study, and at integrating the currently existing apparatus of comparative and reconstructive studies in the domain of meter and rhythm.

The poetic corpus of the NCRL (the National Corpus of the Russian Language) is not intended for the study of rhythm-shaping elements of verse. The corpus does not include an apparatus for extracting summary statistical data pertaining to the chief criteria for the analysis of metrical texts. Moreover, this corpus, as well as the unique system devised by Pil’shchikov, only allows for research on Russian language material.

³ See Liubov’ Zlatoustova [6].

Skulacheva's studies attempt to analyze cognitive processes that take place during the perception of rhythmic texts that are read aloud.

The CSAPPT will make it possible to create reconstructions based on written sources by relying on the RS-theory of Marina Krasnoperova. Within this theory, Kolmogorov's probability language model of verse meter received a cognitive interpretation; moreover, new probability models were created, based on the vocabulary of verse as well as the vocabulary of prose. The models' correspondence or lack of correspondence to verse parameters provides information on the mechanism of versification, on the one hand, and on the pool of rhythmic elements used by the poet, on the other.

The CSAPPT represents a qualitatively new stage in the development of the comparative study of verse; it summarizes the previous experience in metrics and embraces a cognitive approach. A portion of RS-theory was integrated into CSAPPT – namely the apparatus of language probability models. This new computer system is unique, and now it is being adapted for the study of verse rhythm in different languages.

3 Computer Models of Verse Prosody

Verse prosody is being studied against the background of different computer models of versification. In the RS-theory there are two groups of such models –cognitive and probability. The cognitive models (CM) describe a hypothetical complex of these processes and mechanisms of versification. The probability models (PM) provide a connective link between CM and the texts. The degree to which the PM does or does not conform with the metrical text (verse) allows one to hypothesize about the mechanism of versification. The PM are created by using natural accents in the language on the basis of the prose rhythm (rhythmical vocabularies) and the specific technique of versification. The various models correspond to the varying conditions in which versification occurs.

The independence probability model, or Andrey Kolmogorov's model (IM) [7], corresponds to the freest type of versification and correlates with a fairly advanced stage of verse evolution along with a developed language system of meter. This model assumes that the poet, in trying to observe the meter, almost entirely submits to the nature of the language. In this model, the choice of rhythmical/phonetical word types is carried out in accordance with the principle of independence: i.e., it is random. The other type of probability model is based on the principle of dependence in the choice of rhythmical words, which is determined by their position in a verse line and the preceding rhythmic context; these are the Models of Dependence (MD). Depending on the extent to which these models conform, or do not conform, to the verse parameters, we can create hypothetical reconstructions of the versification mechanisms.

For these reasons, since different types of versification can be assigned to these models, this modelling method – which can provide information about the processes leading to the formation of verse and its rhythmic pattern – has been employed in comparative prosody. The first experiments in applying this method to the comparative study of verse have been provided by the author of this paper [8, 9].

This method has now been computerized and built into the CSAPPT, designed for computational mechanisms that would make it possible to create a general statistical analysis of verse and prose rhythm and to reconstruct the processes of metrical verse generation. This task opens new perspectives for research on the rhythmic properties of speech and on the communication of semantic signals by means of prosodic shapes of accentual complexes. This is pertinent to the development of artificial intelligence in the domain of speech rhythm.

The CSAPPT is an analytical system that includes large corpora of texts in different languages: English, German, Dutch, Russian, Ukrainian, etc. A great emphasis in CSAPPT is placed on the rhythmic interpretation of metrical and non-metrical texts.⁴ Using this system requires a rhythmic dictionary that contains statistical data on the distribution of rhythmical words in prose or in verse. A rhythmic word is a complex of syllables united by one accent, like *by-daý*, *my-father*, and *and-tomórow*. Every rhythmical word has a special number in the prosodic vocabulary, for example, 1.1. (monosyllabic), 2.1. (disyllabic with stress on the first syllable), 2.2. (on the second), 3.1. (trisyllabic with an accent at the beginning), etc., as well as a number indicating the frequency of the word's usage by one or another author, in a particular text or language, see Table 1.

Table 1. Prosodic Vocabularies of Prose in Different Languages

Word	German	Russian			Ukrainian
		Lomonosov	Trediakovsky	Pushkin	
1.1.	0,204	0,052	0,070	0,080	0,072
2.1.	0,258	0,112	0,130	0,112	0,161
2.2.	0,178	0,133	0,232	0,168	0,196
3.1.	0,028	0,066	0,057	0,061	0,055
3.2.	0,172	0,152	0,127	0,156	0,186
3.3.	0,044	0,040	0,094	0,097	0,091
4.1.	0,006	0,013	0,007	0,008	0,007
4.2.	0,026	0,118	0,087	0,085	0,061
4.3.	0,032	0,088	0,065	0,098	0,088
4.4.	0,015	0,009	0,029	0,021	0,020
5.1.	0,001	0,002	0,003	0,001	0,000
5.2.	0,002	0,018	0,011	0,012	0,007
5.3.	0,015	0,094	0,041	0,042	0,025
5.4.	0,008	0,022	0,019	0,025	0,020
5.5.	0,002	0,001	0,003	0,003	0,001
6	0,005	0,066	0,019	0,025	0,010
7	0,002	0,015	0,006	0,005	0,000
8	0,002	0,000	0,002	0,000	0,000

This table presents the distribution data for the main rhythmical words in German, Russian and Ukrainian. The rhythmical vocabularies of individual languages differ; German prose, where the frequency of monosyllabic and disyllabic words is significantly higher,

⁴ When forming a system of rules for the rhythmic presentation of verse and prose, both our developments and the developments of leading specialists are considered [10–12].

stands out. However, there may be differences within the same language due to differences in time and writing style, see Fig. 1.

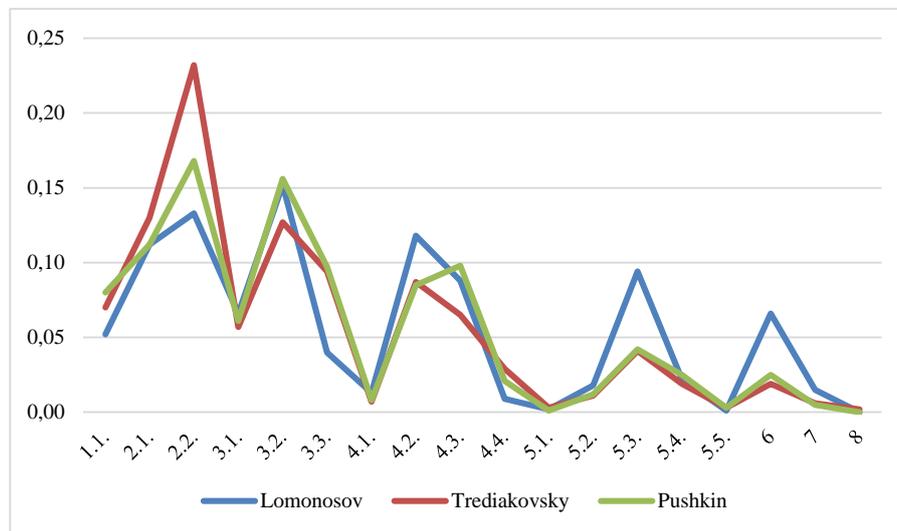


Fig. 1. Distribution of Rhythmical Words in Russian Prose

Pushkin's rhythmical vocabulary is close to the average and, on the whole, turns out to be more similar to the rhythm of the simple language in Trediakovsky's novel *Riding to the Island of Love*, while the rhythm of the prose in Lomonosov's high style is distinctive. The question arises, what rhythmic vocabulary do poets use in creating verse, from what linguistic reservoir do they take their prosodic material?

The CSAPPT allows us to study the rhythm of verse against the background of prose prosody, instantly produce statistics on the distribution of rhythmic structures in verse, calculate stress profiles, and visualize the data. The new system has opened a path for the computational study of the unconscious processes reflected in verse rhythm. An analysis of verse employing this methodology enables us to reconstruct the mechanisms for the development of metrical texts and to investigate the role of language in this process. This system also makes it possible to analyze what linguistic stock of prosody poets use and from what source they draw music for verse. The statistical data on the deviations of stress patterns from the meter may constitute the «prosodic portrait» of a poet or of a whole epoch.

In the rhythmical analysis special attention will be paid to the most significant parameter of verse prosody, namely, the frequency of word accents on the metrically strong positions (SP). The statistics for the stressing on different SP appears in the stress profile of verse. In cognitive poetics, this profile is conceptualized as a perceptive vision of the overall prosodic picture of a poem, see Table 2 and Fig. 2.

Table 2. Data of Stress Profiles of Russian Verse

	I	II	III	IV
End of 18 th c. (Derzhavin)	0,844	0,750	0,500	1,000
19 th c. (Lermontov)	0,825	0,956	0,410	1,000
20 th c. (Ivanov)	0,845	0,842	0,451	1,000

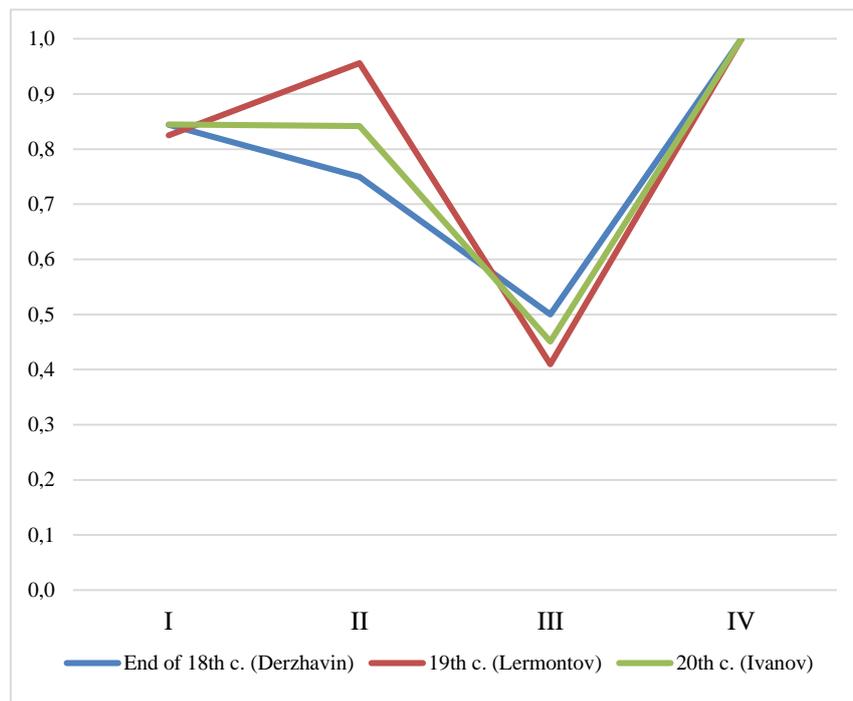
**Fig. 2.** Stress Profiles for the Russian Iambic Tetrameter

Fig. 2 shows how the rhythm of the Russian iambic tetrameter changes over time. From the end of the 18th century (Derzhavin) to the 19th century (Lermontov), the main transformation we can observe is that the prosodic picture for the first half of the verse line is greatly transformed: the second S-position becomes clearly stronger than the first, and the iambic rhythm assumes an alternating form. In the 20th century (Vyacheslav Ivanov), there is a weakening of the alternation – a reverse tendency arises, the rhythm becomes similar to that of the late 18th century.

4 The Study of Verse Prosody Using Probability Models

In previous works by the author of this paper, as well as in studies by Krasnoperova, it was shown that, in general, for most literary traditions, the dependency model is best suited to describe the early stage of metrical verse evolution, but the independence

model better matches the advanced mechanism of versification. In addition, it should be noted that the DM assumes that the poet devotes significantly more effort to generating a metrical line than in the case when the versification is carried out in accordance with the conditions of the IM.

In the DM the value for the creative effort involved in the search for an appropriate rhythmical word at a given place in the metrical line equals the probability that success will occur no later than on the n -th attempt:

$$\sum_{k=0}^{n-1} f(k, 1, p), \quad (1)$$

where k is the number of unsuccessful attempts, p is the probability of success, and $f(k, 1, p)$ is the probability of success on the $k+1$ -th attempt.

In the IM this value is divided by a certain standard quantity $f(0, 1, p)$, namely, the probability of success being achieved on the first attempt (i.e. the word appropriate to the metre is chosen on the first attempt). Thus, in the independence model for each rhythmical word, effort is evaluated according to the formula:

$$\frac{\sum_{k=0}^{n-1} f(k, 1, p)}{f(0, 1, p)} \quad (2)$$

(the probability of success no later than on the n -th attempt divided by the probability of success on the first attempt).

Thus, the more mature manner for generating verse rhythm corresponds to the IM — in the poet's mind, an internal standard for measuring the amount of effort spent is formed. The dependence model assumes the situation when a verse line will be generated in any case; in accordance with the independence model, a metrical verse line may not always be formed. In other words, the DM describes versification taken to the bitter end — a metrical line will be created at any price.⁵

By dint of CSAPPT we want to analyze what type of probability model is close to one or another period of versification, to one or another literary tradition. The model can also serve as an important unconscious guide for the author in the formation of a metrical text. Previous studies have shown that often in a certain period different authors develop a common stress profile of iambic verse. The view that the poetry is a self-organizing system made it possible to put forward various hypotheses about the typology of versification mechanisms based on data for the stress profile in diachrony. However, despite the general laws that influence the formation of one profile or another, certain authors have their own specific profile, which some researchers interpret as the individual rhythmic portrait of a poet. So, for example, in the early stages of the development of Russian verse, Lomonosov and Sumarokov develop similar but fundamentally different forms of stress profiles.

The stress profile of Lomonosov's verse differs from that of the late period of the Russian tetrameter (see Fig. 2), while the rhythm of Sumarokov's iamb is close to that of Derzhavin's verse (see Fig. 3). The profile diagrams for both poets assume the form of a clear frame structure, in which the external S-positions, the first and fourth, are the

⁵ For more details see Krasnoperova, Kazartsev [13].

most stressed; however, the distribution of stresses on the internal S-positions is somewhat different. Lomonosov has a tendency towards strong and almost equal stressing of the second and third SP. In Sumarokov, the third SP is significantly weaker than the second. It turns out that these two stress profiles are predicted by different types of probability models. Thus, the model of independence predicts a stress profile close to that found in Sumarokov's versification, but Lomonosov's verse rhythm is better described by the model of dependence (Table 3, Fig. 3).

The data produced by the models in both cases are very close to those of the verse. Recall that the model of dependence corresponds to a more arduous versification, "to the bitter end," where the amount of effort spent on generating a poetic line is not normalized. On the contrary, the model of independence is matched with a type of versification, in which the poet has formed a certain standard of effort devoted to generating the verse line. In accordance with these data, it turns out that the versification of the early Lomonosov is very constrained; he strives to create an iambic verse at all costs. As a result the verse turns out to be very heavily stressed, too iambic. But Sumarokov, even in his early poems, strives for freedom, not following the policy of using any means to generate pure iambic verse.

Table 3. Data of Stress Profiles of Early Russian Iambs and Probability Models

	I	II	III	IV
Early Lomonosov	0,979	0,900	0,871	1,000
Early Sumarokov	0,939	0,783	0,589	1,000
Model of Dependence	0,955	0,914	0,881	1,000
Model of Independence	0,979	0,767	0,565	1,000

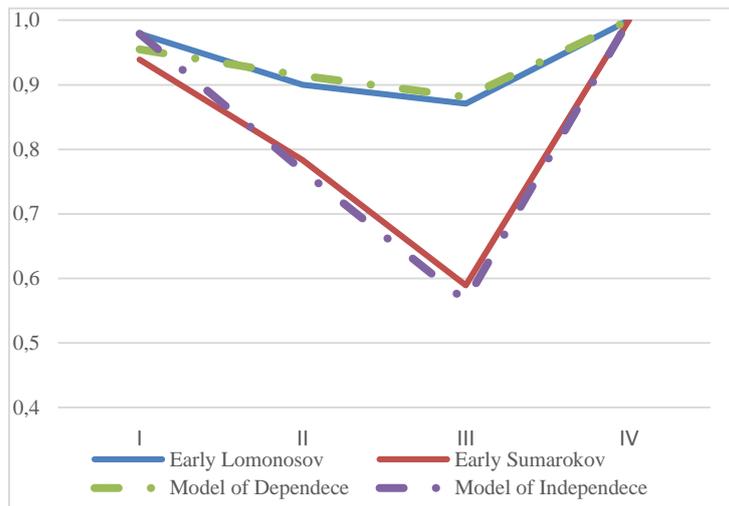


Fig. 3. Stress Profiles of Early Russian Iambs Against the Background of Probability Models

A more detailed comparison of the early iambic tetrameter in German, Russian and Ukrainian poetry gives a completely new picture regarding the possibilities for computer simulation of verse prosody. The data make it possible to see that language models of dependence can predict not only the frame profile for the distribution of stresses, but also the alternating profile, which is characteristic of a certain stage in the development of early iambic forms in German and Ukrainian metrical poetry. The alternating stress profile is the result of the regressive accentual dissimilation process in verse lines. This tendency proceeds from the end of the verse line to the beginning of it. As a result, the last S-position becomes the most striking (the most iambic), the penultimate position becomes the least iambic, the second SP in iambic tetrameter is again strengthened, and the first SP is weakened, compared to the second. Such a distribution of stresses, as we see, develops in some of the early iambic poems of German and Ukrainian poetry, and it is fairly well described by the corresponding computer models of dependence created from dictionaries of German and Ukrainian poetry (see Table 4 and Fig. 4).

Table 4. Data of Stress Profiles of Early German and Ukrainian Iambs and Probability Models⁶

	I	II	III	IV
German Iamb (Gryphius)	0,906	0,913	0,888	0,978
Ukrainian Iamb (Shevchenko)	0,797	0,902	0,257	1,000
Model of Dependence (German)	0,921	0,913	0,880	0,973
Model of Dependence (Ukrainian)	0,705	0,820	0,702	1,000

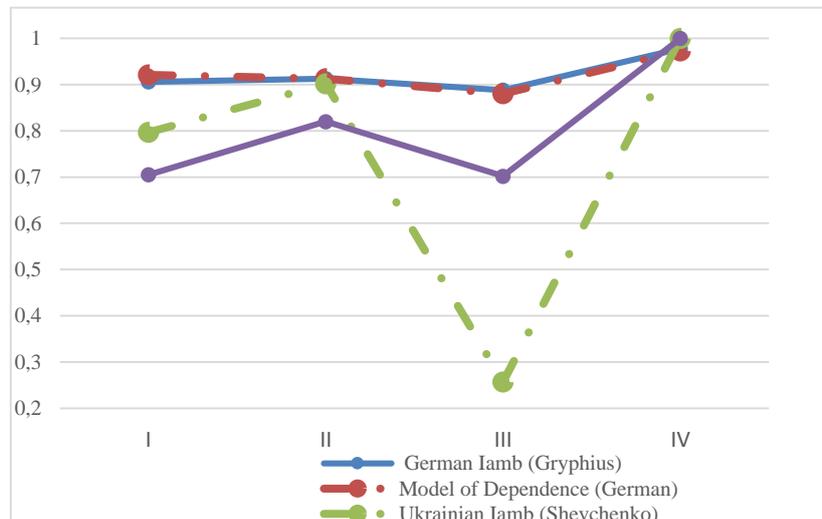


Fig. 4. Stress Profiles of Early German and Ukrainian Iambs Shown Alongside Probability Models

⁶ Data of Ukrainian verse and model are from the research of my student Arina Davydova.

The table and figure show that in Ukrainian verse the alternating tendency is more pronounced than in the German. However, in both cases, the two models of dependence (the one is built on the rhythm of German language parameters of prose, the other on the Ukrainian language) demonstrate a similarity to the verse. For the early Russian iamb, this type of stress profile is not characteristic. We observe a clear alternation in the 30-40s of the 19th century, as in Lermontov, (see Fig. 2). However, in the Russian 18th century—in the verse of Lomonosov, Sumarokov and Derzhavin—the frame stress profile is preserved in its pure form, which is similarly derived from premises of the Russian language. Note that the earliest iambics of Lomonosov are described by the dependence model, but starting with Sumarokov, the mechanism of versification obviously changes, the verse better corresponds to the model of independence with its freer technique of versification. At the same time, for the classical forms of German and Ukrainian iambic verse, the more rigid manner of generating verse, corresponding to the model of dependence, obviously remains relevant. However, despite the fact that the same model is used to predict the versification of Lomonosov, as well as that of German and Ukrainian verse, the characters of the predicted rhythm are completely different: with the Russian material, DM predicts a frame profile for the iambic tetrameter, but with the German and Ukrainian material it predicts an alternating profile. This is due to the difference in the language material.

5 Conclusion

It turns out that while all three traditions maintain the same type of versification in the development of metric verse, language features determine the appearance of alternation. Thus, the linguistic model of dependence, along with the versification technique that corresponds to it, allows one to predict not only the frame structure, but also the alternating rhythm of iambic verse.

This study as a whole allows us to say that in the analysis of various probability models for verse, the differences between them can be caused not only by the structure of the model, based on one or another type of versification, but also by the language rhythm. The influence of language in this case can be so significant that it changes the model predictions. Therefore, when analyzing modeled structures, not only is the cognitive aspect in the process of verse formation important, but so is the material that is involved in this process.

The computer modeling of verse rhythm, based on certain principles of verse generation, allows us to identify and describe the features of versification under certain conditions in a particular language and to study the typology and differences in the implementation of the same verse meter in different traditions. Using this technique, we were able to perceive the difference in the mechanisms of versification by the two founders of Russian metric verse, Lomonosov and Sumarokov, and we were able to explain what caused the great rhythmic freedom of Sumarokov's verse and the way it differs from Lomonosov. This is probably due to the difference in the amount of effort devoted to the generation of strictly iambic verse by both poets.

This research also showed that, despite the similarities in the mechanisms of versification, the rhythm of verse in individual traditions can be different. There is reason to believe that in German, Russian and Ukrainian verse at certain times the same versification mechanism operates and is described by the model of dependence, but the rhythm of the verse is not similar: in Russian iambic verse stressing on the S-positions forms a frame; in German and Ukrainian it creates an alternating rhythm.

References

1. Zhirmunsky, V.M.: O nacional'nyh formah jambicheskogo stiha [About the national forms of Russian iambic verse]. *Teorija stiha* [Theory of verse], 7–23. Nauka, Leningrad (1968).
2. Krasnoperova, M.: *Osnovy rekonstruktivnogo modelirovanija stihoslozhenija: na materiale ritmiki russkogo stiha* [The basics for reconstructive simulation of versification, based on the rhythm of Russian verse]. Saint Petersburg State University Press, St. Petersburg (2000).
3. Krasnoperova, M.: Cognitive aspects of probabilistic-statistical analysis in reconstructive simulation of versification. In: *Cognitive Modeling in Linguistics. Proceedings of 10th International Conference*, 110–122. Becici – Kazan (2008).
4. Pilshchikov, I.: Rhythmically Ambiguous Words or Rhythmically Ambiguous Lines? In *Search of New Approaches to an Analysis of the Rhythmical Varieties of Syllabic-Accentual Meters*. In: *Quantitative Approaches to Versification*, 193–200. The Institute of Czech Literature of the Czech Academy of Sciences, Prague (2019).
5. Skulacheva, T.: Verse and Prose: A Linguistic Approach. In: *Poetry and Poetics: A Centennial Tribute to Kiril Taranovsky*, 239–248. Slavica Publishers, Bloomington, Indiana (2014).
6. Zlatoustova, L.: O novom napravlenii v stihovedenii: Teorija rekonstruktivnogo modelirovanija stihoslozhenija [About a new method in verse study: the theory of reconstructive simulation]. *Moscow State University Bulletin: Philology*, 5, 122–125. Moscow State University, Moscow (2004).
7. Kolmogorov, A.: Primer izuchenija metra i ego ritmicheskikh variantov [An example of studying of meter and its rhythmic variants]. *Teorija stiha* [Theory of verse], 145–167. Nauka, Leningrad (1968).
8. Kazartsev, E.: Comparative Study of Verse: Language Probability Models. *Style* 48(2), 119–139. Penn State University Press, Pennsylvania (2014).
9. Kazartsev, E.: Quantitative und vergleichende Versforschung – Ausweg aus der Krise?. *Jahrbuch für Internationale Germanistik*, 48(1), 53–72. Peter Lang, Bern (2016).
10. Kazartsev, E.: Ritmicheskoe predstavlenie teksta v sravnitel'nyh stihovedcheskih issledovanijah [Conceptions of rhythmical structure in comparative studies of verse]. *Russian Linguistics* 42(2), 271–287. Springer, Heidelberg (2018).
11. Tarlinskaja, M.: Fletcher's Versification. *Studia Metrica et Poetica* 7(1), 7–33. University of Tartu Press, Tartu (2020).
12. Zhirmunsky, V.: *Introduction to Metrics: The Theory of Verse*. De Gruyter Mouton, London (1966).
13. Krasnoperova, M., Kazartsev, E.: Reconstructive simulation of versification in comparative studies of texts in different languages (theoretical aspects and practice of application). *Frontiers in Comparative Metrics*, 97–120. Peter Lang, Bern (2011).