

Automated Rhythmic Device Search in Literary Texts Applied to Comparing Original and Translated Texts as Exemplified by English to Russian Translations

N. S. Lagutina^{a, *}, K. V. Lagutina^{a, **}, E. I. Boichuk^{b, ***},
I. A. Vorontsova^{b, ****}, and I. V. Paramonov^{a, *****}

^a*P.G. Demidov Yaroslavl State University, Yaroslavl, 150003 Russia*

^b*Yaroslavl State Pedagogical University named after K.D. Ushinskii, Yaroslavl, 150000 Russia*

**e-mail: lagutinans@rambler.ru*

***e-mail: lagutinakv@mail.ru*

****e-mail: elena-boychouk@rambler.ru*

*****e-mail: arinna1@yandex.ru*

******e-mail: Ilya.Paramonov@fruct.org*

Received August 12, 2019; revised September 11, 2019; accepted September 13, 2019

Abstract—Analysis of a translated text’s functional equivalence to its original based on the attainment of rhythmic equivalence is an extremely important task of modern linguistics. The rhythmic component is an integral part of functional equivalence, which cannot be attained without reproducing the rhythmic devices of the original text. For the purposes of analyzing rhythmic devices in an original literary text and its translations, the authors have developed the software tool ProseRhythmDetector, which can find and visualize lexical and syntactic devices in English- and Russian-language prose, specifically anaphora, epiphora, symploce, anadiplosis, epanalepsis, reduplication, epistrophe, polysyndeton, and aposiopesis. The objective of this study is to present the results of the testing of ProseRhythmDetector on two English-language literary texts and their Russian translations: “Villette” by Charlotte Brontë and “The Black Prince” by Iris Murdoch. Using the results obtained by the software tool, the authors cross-referenced rhythmic devices in the original and translated texts and compared rhythm aspects and their contexts. The experiment made it possible to identify specific characteristics of a translator’s rendition of the author’s style, as well as detect and explain cases in which the use of rhythmic devices in the translated texts do not correspond to the originals. The application of the ProseRhythmDetector software tool significantly reduced the researchers’ workload by automatically retrieving lexical and syntactic devices with a rather high accuracy (from 62% to 93% for different rhythmic devices).

Keywords: prose rhythm, rhythm analysis, natural language processing, rhythmic devices, automation

DOI: 10.3103/S0146411620070147

INTRODUCTION

One of the problems of modern linguistics is the task of determining and analyzing the rhythm of literary texts. Rhythm is usually understood as a regular repeated pattern of similar and comparable units of natural language that perform structuring, text-forming, and expressive-emotional functions. Rhythmic devices of prose and poetry have independent characteristics and are not only a phenomenon of organization and perception of literary text, but also contribute to defining the author’s style [1].

Research in the field of automatic rhythm analysis is mainly concerned with poetic texts, for which such analysis is based on identifying patterns of alternation of stressed and unstressed syllables and simple quantitative characteristics such as sentence length and word frequency. However, rhythmic devices include a large number of aspects at the phonetic, lexical, morphological, and syntactic levels. In this regard, an interesting and difficult task is to create methods and tools for determining rhythm and describing it in specific parameters. With such parameters, researchers could compare rhythms of texts by different authors, of different parts of one text, or of an original text and its translation [2]. Linguists also note

the importance of preserving the author's rhythm when translating a literary text into another language and the fact that that rhythm is frequently distorted in translations of prose texts [3].

The research on rhythm in literary translation is complicated by the extreme complexity of the object of research –prose rhythm, the intricacy and multidimensionality of the phenomenon of literary translation, as well as the limited mathematical tools and the lack of effective automated systems for evaluating the parameters of the rhythm of the original and translated texts. However, some researchers do use mathematical methods to evaluate rhythmic devices [4].

With all of the above in mind, the authors set the objective of constructing a system of automated retrieval of the complex of rhythmic devices used in literary texts and applying it to comparative analysis of English-language original texts and their Russian translations. An important feature of the developed tool is that it visualizes the rhythmic device search results for a given text, making expert analysis of individual aspects easier.

CONSIDERED RHYTHMIC DEVICES

In this study, the rhythmic characteristics of a text are determined by examining the use of rhythm-creating literary devices, which are based on repetition of a certain number of elements in a certain configuration and in a certain position. Let us define the following rhythmic devices:

1. Anaphora – a way of connecting text segments (parts of a phrase, lines of poetry) by repeating a word or phrase in the starting position of these segments.

2. Epiphora – a way of connecting text segments (parts of a phrase, lines of poetry) by repeating a word or phrase in the final position of these segments.

3. Symploce – a device of syntactic parallelism in adjacent phrases or lines of poetry that have the same beginning and the same end but different middle parts, or vice versa, different beginnings and ends but the same middle part.

4. Anadiplosis – rhetorical device in which a sentence begins with the same words that the previous sentence ends with.

5. Epanalepsis – a figure of speech that consists of repeating the same word or phrase with slight variations.

6. Reduplication – a figure of speech that consists of repeating the same word or expression twice.

7. Epistrophe – a stylistic device that consists of repeating the same word or expression in a long phrase or period.

8. Polysyndeton – a stylistic device that consists of deliberately increasing the number of conjunctions in a sentence, usually between coordinate elements.

9. Aposiopesis – an omission, incompleteness of thought, suddenly breaking off in oral or written speech, most often in the form of interrupting a sentence and beginning a new one.

Rhythm analysis is performed in this study on the basis of automated search for these rhythmic devices and subsequent quantitative processing of search results. These specific devices were chosen due to their high frequency in prose texts and the fact that most linguistic research in the field of rhythm analysis deals with these rhythmic devices at the lexical and grammatical level.

OVERVIEW OF RELATED RESEARCH

Rhythmic properties of a text are most often determined on the basis of phonetic aspects of rhythmicity, which are based on a structural analysis of the text at the level of syllables, on its metro-rhythmic parameters, and on identifying rhymes. This approach is used in the existing software tools, such as SPARSAR¹ for the English language [6], Metricalizer² for German [7], Rhymes³ and RitmInMe⁴ for Russian. All these tools are meant for poetic texts.

A similar approach is often applied to prose as well. The phonetic rhythms of English-, Romanian-, and French-language texts are analyzed in a series of studies [8, 9]; the results of the analysis are applied to identifying the style of the texts (poetry, literary prose, or political speech). Similar research has been done on Russian-language texts [10]. E.I. Boichuk et al. [11] analyze the rhythm of French literary prose

¹ <https://sparsar.wordpress.com>

² <https://www.poetron-zone.de/metricalizer.php>

³ <http://rifmovnik.ru>

⁴ <http://www.ritminme.ru>

in order to define the authors' styles. The authors of [12] transform the text into a set of signals by splitting it into letters and then investigate that text's rhythm using mathematical methods of decomposing signals into functions (Empirical Mode Decomposition). This approach analyzes texts at the level of syllables, letters, and sounds—less often, words—and conveniently allows easily used statistical tools to calculate frequency, recurrence rate, correlation, etc. However, an in-depth study of prose rhythm requires working at a higher semantic level of phrases (syntagms), sentences, and paragraphs.

Researchers in the field of computational linguistics consider it necessary to study various speech figures of natural language and note the difficulty of formalizing definitions of such figures and of constructing methods for automated search for them [13]. The authors of this article have described the process of creating an ontology and a database of metaphors, rhetorical questions, juxtapositions, and figures of speech that are based on repetition of words and phrases. Special attention was drawn to the fact that the interaction of linguists and information technology specialists opens up possibilities to achieve a deeper understanding of the structure and functioning of human language.

Unfortunately, an insufficient number of studies on automating identification and application of lexical and syntactic aspects of text rhythmicity exist at this time. The authors of [14] propose a method for detecting anadiplosis, anaphora, antimetabole (repetition of the words of the preceding phrase, but in a different order and with different meaning), and epiphora using regular expressions. The experiment was performed on a corpus of over 163 thousand text fragments, but only the frequency of each of the considered rhythmic devices was calculated, and the accuracy and completeness of the results were not assessed.

Russian scientists [15] compare six software systems for detecting anaphora. To automate the search process, the systems used machine learning methods, morphological and syntactic parsers, ontologies and rules. In these experiments on a corpus of 85 texts, the best F-measure scores were achieved when linguistic algorithms based on rules and ontologies were used.

The study of methods of retrieving chiasmus (crosswise changes the order of elements in two parallel sets of words), epiphora, and epanaphora (repeated first and final words in adjacent phrases or sentences) is described in detail in [16]. The authors note the difficulty of defining and formalizing parameters of a text and its components as is necessary for a high-quality figures of speech search. A superficial analysis of the structure of a text returns too many "false" results—fragments that contain repeated elements, but do not correspond to the relevant rhythmic device in meaning, which happens especially often with chiasmus. Uses of rhythmic devices can be very rare, for example, one instance per several hundred pages of text. The authors chose a linear model for selecting candidates and used it to develop manually configurable detectors. When a sufficiently large training sample was accumulated, the system was trained using logistic regression. As a result, the value of the F-measure in most cases exceeded 60%. The authors position their tool as a linguist's assistant, complementing the traditional manual analysis for the conditions of working with large numbers of literary texts.

To sum up, the construction and study of automatic text analysis algorithms at the level of semantic structures, such as rhythmic devices, is an important and understudied task of computational linguistics. A part of solving this problem is developing tools that can assist linguists. These tools are necessary for experiments with large text corpora, for which manual processing is not realistic. In this case, the research results include both the parameters for evaluating texts and materials that can be used in machine learning.

RHYTHMIC DEVICE SEARCH ALGORITHMS

For the tasks of analyzing the rhythm of a literary prose text and comparing it with the rhythm of its translated version, a set of algorithms that automatically retrieve rhythmic devices (specifically lexical and syntactic ones) from a text was developed.

The input data for all algorithms is raw text. A text T consists of n sentences $\langle s_1, s_2, \dots, s_n \rangle$, each sentence is a set of words $\langle w_1, w_2, \dots, w_m \rangle$. Each algorithm returns a list of aspects A of a given rhythmic device: A_{anph} for anaphora, A_{eph} for epiphora, A_{sym} for symproce, A_{andp} for anadiplosis, A_{epnl} for epanalepsis, A_{redpl} for reduplication, A_{epstr} for epistrophe, A_{poly} for polysyndeton, and A_{aps} for aposiopesis.

Each aspect $a = \langle W, S \rangle$, $a \in A$ consists of a word or phrase W repeated in several sentences $S = \langle s \mid s \in T, W \in s \rangle$, $|S| > 1$. The list of sentences S is the context for the aspect a . Researchers can use the contexts to compare the original and translated texts when analyzing how the translator reproduces the rhythmic devices used by the original author. The context can also be used to verify that the aspect was identified correctly.

Let us provide an example of an occurrence of a rhythmic device in a text and the associated notation: *He does. He is fond of you. You are his favourite.*

This is an example of anadiplosis. Its presence is considered as an aspect $a \in A_{andp}$. $a = \langle W, S \rangle$, where the repeated words $W = \langle \text{you} \rangle$ and the context $S = \langle \text{“He is fond of you”, “You are his favourite”} \rangle$.

Let us consider the developed rhythmic device search algorithms in more detail.

Anaphora, epiphora, and symploce search algorithms

The first step of these algorithms is creating a list of anaphora candidates: $K = \langle k_1, \dots, k_l \rangle$, where each $k \in K$ has the structure $\langle w, S' \rangle$ and consists of a word w and a list of adjacent sentences $S' = \langle s \mid s \in T, |s| > 1, s = \langle w, w_1, w_2, \dots, w_m \rangle \rangle$.

An aspect $a = \langle W, S \rangle$ is constructed for each k . It is constructed using a loop: initially $W = \langle w \rangle$, then at each j -th step another word w_j such that $w_j \in s$ for several sentences $s \in S'$ is added to W . If several such w_j are found for different sentences, the w_j corresponding to the largest number of sentences is chosen. After w_j is added to W , S' is filtered so that only those s that satisfy the condition:

$$s = \langle w, w_1, w_2, \dots, w_j, w_{j+1}, \dots, w_m \rangle,$$

i.e., sentences that start with a set of words W but are not limited to just that set of words, remain. When no suitable w_j can be found, the loop ends, $a_i = \langle W, S \rangle$, where W is the list of repeated words, and $S = S'$ after the algorithm stops.

The epiphora search is performed in a similar way, the only difference being that W is formed from the end of the sentence instead of the beginning.

The list of symploces A_{sym} is compiled from the lists of anaphoras and epiphoras, A_{anph} and A_{eph} respectively. All pairs of aspects $\langle b, c \rangle$ such that $b \in A_{anph}$, $c \in A_{eph}$ are compared to each other. If $b = c$, that means that these sentences of the same context have the same beginning and the same ending, so a symploce has been found. Then b is removed from A_{anph} , c – from A_{eph} , and b is added to A_{sym} .

The algorithm stops as soon as it has gone over all pairs $\langle b, c \rangle$.

Anadiplosis search algorithm

Let us represent the analyzed text as a set of words w and punctuation marks p : $T = \langle w_1, p_1, w_2, \dots, p_q \rangle$. The algorithm searches for uses of anadiplosis using a loop that identifies aspects with a greater $|W|$ at each step.

At the first step, the anadiplosis search algorithm goes over the text searching for lists with the structure $l = \langle w, p, w \rangle$, i.e., two identical words separated with a punctuation mark. On the basis of l , the aspect $a = \langle W, S \rangle$ is constructed, where $W = \langle w \rangle$, $S = \langle s \rangle$, and $l \in s$ or $S = \langle s_1, s_2 \rangle$, $s_1 = \langle w_1, \dots, w \rangle$, $s_2 = \langle w, w_{21}, \dots, w_{2m} \rangle$, if p signifies the end of the sentence.

The next step is almost completely identical to the previous one, except for the fact that $l = \langle w_1, w_2, p, w_1, w_2 \rangle$; i.e., the search is now for *pairs* of identical words separated with a punctuation mark. Then $W = \langle w_1, w_2 \rangle$.

Generally, at the i -th step the algorithm searches for a list with the structure

$$l = \langle w_1, w_2, \dots, w_i, p, w_1, w_2, \dots, w_i \rangle,$$

and $W = \langle w_1, w_2, \dots, w_i \rangle$. The algorithm stops after step number $i_{\max} = \max_{s \in T} |s|$.

Epanalepsis, reduplication, and epistrophe search algorithms

The reduplication and epistrophe search algorithms go over all words $w \in T$. A search for all sets of consecutive words $W = \langle w_1, \dots, w_k \rangle$, $k > 0$ that form the list $l = \langle w_1, \dots, w_k, \dots, w_1, \dots, w_k \rangle$, $l \in T$, i.e., W

occurs in T several times, with the condition that W cannot consist of only prepositions or conjunctions, is performed.

If a list with the structure $l = \langle w_1, \dots, w_k, w_1, \dots, w_k \rangle$, i.e., W is doubled while $W \in s$ (W is doubled within a single sentence), and for any $a' \in A_{andp}$, $W \notin a'$ (not anadiplosis), is found, that is a reduplication. Then an aspect $a = \langle W, S \rangle$, where $S = \langle s \rangle$, is formed, and a is added to A_{redpl} .

If a list with the structure $l = \langle w_1, \dots, w_k, w, \dots, w_1, \dots, w_k \rangle$, where $l = \langle s \rangle$ or $l = \langle s_i, s_{i+1} \rangle$, $i = 1, \dots, n - 1$, i.e., W is repeated in the beginnings and in the ends of a sentence or a pair of sentences and, possibly, in their middle parts, is found, that is an epistrophe. Then an aspect $a = \langle W, S \rangle$, where $S = l$, is formed, and a is added to A_{epstr} .

If a list with the structure $l = \langle w_1, \dots, w_k, w, \dots, w_1, \dots, w_k \rangle$, i.e., at least one word w is present between the repeated W , while $W \in s$, $W \in s_{i+1}$, $i = 1 \dots n - 1$, and also for any $a' \in A_{anph}, A_{eph}, A_{sym}, A_{andp}, A_{epstr}$, $W \notin a'$, that is an epanalepsis. Then an aspect $a = \langle W, S \rangle$, where $S = \langle s \rangle$ or $S = \langle s_i, s_{i+1} \rangle$, is formed, and a is added to A_{epnl} .

Search algorithms for the syntactic devices (polysyndeton and aposiopesis)

To perform a polysyndeton search, the algorithm needs lists of conjunctions: a list of regular conjunctions $C_1 = \langle \text{but, till, as, if, after, until, because, and, or, nor, so, before, since, that, till, until, unless, whether, while, where, when, why, what, how, whenever, although, though, once, than, whereas, thus} \rangle$ and a list of correlative conjunctions $C_2 = \langle \langle \text{both, and} \rangle, \langle \text{either, or} \rangle, \langle \text{not only, but} \rangle, \langle \text{not only, but also} \rangle, \langle \text{rather, or} \rangle, \langle \text{just as, so} \rangle, \langle \text{neither, nor} \rangle, \langle \text{whether, or} \rangle, \langle \text{if, then} \rangle \rangle$, as well as a list of conjunctive adverbs $C_3 = \langle \text{after all, as a result, for example, in addition, in fact, in other words, on the other hand} \rangle$. Corresponding lists were provided for the Russian language as well.

Each of the lists C_i , where $i = 1, 2, 3$, is processed as follows.

For each conjunction or conjunctive adverb $c \in C_i$ and each sentence $s \in T$, the condition $l \in s$ is checked, where l has one of the following structures:

$$\begin{aligned} l &= \langle w_1, \dots, c, \dots, c, \dots, w_m \rangle, \\ l &= \langle c, \dots, c, \dots, w_m \rangle, \\ l &= \langle w_1, \dots, c, \dots, c \rangle, \\ l &= \langle c, \dots, c \rangle, \end{aligned}$$

i.e., c occurs in s more than once, possibly in the beginning or in the end of the sentence. If the condition is met, the aspect $a = \langle W, S \rangle$, where $W = \langle c \rangle$, $S = \langle s \rangle$, is added to the list of aspects A_{poly} .

The aposiopesis search algorithm returns a list of aspects A_{aps} , where each aspect $a \in A_{aps}$ consists of a single sentence $s \in T$. It is the simplest of all the rhythmic device search algorithms, as it consists of just searching for $s \in T$ that contain ellipses and adding $a = \langle \emptyset, s \rangle$ to the list of aspects A_{aps} . These aspects do not contain repeating words.

Software implementation of the algorithms

The above rhythmic device search algorithms are implemented in the ProseRhythmDetector web application, which was developed using HTML, CSS, and JavaScript and published on GitHub⁵.

The users can upload literary texts in Russian or English, and the application will perform a search for rhythmic devices in the uploaded text using the above algorithms. The application then displays the text and highlights the retrieved aspects and their contexts with different colors. In addition, ProseRhythm-Detector calculates the frequencies of the rhythmic devices in the text and can display a list of aspects of a specified device on a separate panel. The main page of the application with a processed chapter of the translation of Charlotte Brontë's novel "Villette" is shown in Fig. 1.

⁵ <https://github.com/text-processing/html-tool>

Fig 1. The main page of the ProseRhythmDetector application.

The application can be used to help researchers evaluate the style of a text as a whole by viewing and analyzing the lexical and syntactic devices found in it. Besides that, using the application in experiments that involve studying features of several literary texts renders these experiments much less time-consuming than studying the texts manually, without automated tools.

EXPERIMENTS

The objective of the experiments carried out in the course of this research was to assess the completeness of reproduction of the above rhythmic devices in the translated texts. The experiments were conducted on two novels: “Villette” by Charlotte Brontë and its Russian translation by N. Mikhal’skaya and “The Black Prince” by Iris Murdoch and its translation by I. Bernstein and A. Polivanova.

In the course of the experiments, the accuracy of the rhythmic device searches performed by the tool ranged from 62% to 93%. Tables 1 and 2 present quantitative results of rhythmic device searches and the number of correspondences between the analyzed literary texts and their translated versions. The retrieved rhythmic devices are discussed in more detail and their uses in the original and translated texts are compared further.

Anaphora

In the text of Charlotte Brontë’s novel “Villette,” anaphora is found in 68 contexts. In the translated text, there are 10 cases in which the anaphoric repetitions completely correspond to the original, with no changes in position in the text or in the number of repeated elements. In 8 out of these 10 cases, the anaphora is pronominal (the repeated elements are subject pronouns *I*, *she*, *it*); this type of anaphora is the most accurately reproduced in the translated text. However, in some cases, the pronoun is replaced with a different one and an additional construction is added in preposition to it:

She had a pale face, hair like night, broad strong eyebrows, decided features, and a dark, mutinous, sinister eye: I noted that she sat close by a little door, which door I was well aware, opened into a small closet where books were kept. **She** was standing up for the purpose of conducting her clamour with freer energies.

Я внимательно взгляделась в нее: бледное лицо, иссиня-черные волосы, широкие выразительные брови, резкие черты лица и темные, мятежные, мрачные глаза. **Я** заметила, что она сидит около небольшой двери, которая, как я знала, ведет в маленький чулан, где хранились книги. Ученица встала, дабы получить большую свободу действий.

In the text of “The Black Prince” by Iris Murdoch, anaphora is found in 585 contexts, of which 102 are reproduced in the translated text with varying degrees of quantitative and positional accuracy. In 45

Table 1. The number of rhythmic devices in the original and translated texts of Charlotte Brontë's novel "Villette"

Device	Original	Translation	Difference	Correspondence
Anaphora	68	211	143	10
Epiphora	10	26	16	2
Symploce	0	2	2	0
Anadiplosis	12	27	15	4
Epistrophe	0	24	24	0
Reduplication	0	486	486	0
Epanalepsis	1649	4518	2869	119
Polysyndeton	179	847	668	2
Aposiopesis	3	57	54	0
Total	1921	6198	4277	137

Table 2. The number of rhythmic devices in the original and translated texts of Iris Murdoch's novel "The Black Prince"

Device	Original	Translation	Difference	Correspondence
Anaphora	585	346	239	102
Epiphora	149	53	96	35
Symploce	9	3	6	3
Anadiplosis	37	37	0	2
Epistrophe	78	0	78	0
Reduplication	23	8	15	2
Epanalepsis	3340	2901	439	1001
Polysyndeton	219	366	147	98
Aposiopesis	0	0	0	0
Total	4440	3966	474	1250

contexts, the number and position of the repeated elements correspond to the original text completely, for example:

I have never tried to please at the expense of truth. I have known, for long periods, the torture of a life without self-expression. *Я никогда не стремился к приятности за счет правды. Я знал долгие мучительные полосы жизни без самовыражения.*

In 22 contexts, the position of the anaphora is retained, but the number of repeated elements is smaller by 1–4 out of 4–6, for example:

I waited. I tried to develop a new routine: monotony, out of which value springs. I waited, I listened. *Я ждал. Я снова постарался выработать упорядоченный образ жизни, создать монотонность, из которой рождаются всплески. Я выжидал, вслушивался.*

In 31 contexts, the position of the anaphora is slightly changed (one sentence forward or backward), and in 18 of these the number of repeated elements is changed, usually decreased, for example:

And I was suddenly deeply frightened by the possibility of having my sister on my hands. I simply did not love her enough to be of any use to her, and it seemed wiser to make this plain at once. I waited for about ten minutes, trying to calm and clear my mind, and then went back to the bedroom door. I did not really expect that Priscilla would have got dressed and be ready to leave. I did not know what to do. I felt fear and disgust at the idea of mental breakdown, the semi-deliberate refusal to go on organizing one's life which is regarded with such tolerance in these days. I peered into the room.

Я боялся, как бы сестра не оказалась вдруг у меня на руках. Я просто-напросто не настолько любил ее, чтобы она могла на меня рассчитывать, и, видимо, лучше всего было сказать ей об этом прямо. Переждав минут десять, пока успокоятся мои нервы и прояснится голова, я встал и подошел к двери в спальню. В сущности, я и не надеялся, что застаю Присциллу одетой и готовой к уходу. Что мне делать, я не знал. Мне противна и страшна была сама мысль о “нервном расстройстве”, этом наполовину сознательном уходе от упорядоченной жизни, к которому в наши дни принято относиться с такой терпимостью. Я заглянул в комнату.

In two cases, the number of repeated elements in the translated text is greater than in the original (three in the translated text against two in the original).

In four contexts, the number of elements is smaller and the repetition is no longer consecutive; as a result, the repetition stops being perceived as anaphoric and instead turns into an epanalepsis, for example:

I went out of the room and closed the door quietly behind me. I heard a soft bound and then the key turning in the lock. I went down the stairs feeling very shaken and, yes, she had been right, disgusted.

Я спустился по лестнице с чувством растерянности и — да, она была права — отвращения. За это время стемнело, солнце больше не сияло на улице, и все в доме стало коричневым и холодным. Я вошел в гостиную, где сидели и беседовали Арнольд с Фрэнсисом.

In 483 translations of contexts that contain anaphoric repetition in the original text, that repetition is either lost or replaced with another rhythmic device, usually epanalepsis, for example:

Arnold Baffin wrote too much, too fast. **Arnold Baffin** was really just a talented journalist.

Арнольд Баффин писал слишком много, слишком быстро. По существу, Арнольд Баффин был всего лишь талантливым журналистом.

Epiphora

In the text of Charlotte Brontë's novel "Villette," 10 occurrences of epiphora were found. Three of the contexts have corresponding versions in the translated text. No complete correspondences in terms of position in the sentence and similarity of use were found in the translated text, which is primarily explained by the specifics of the languages' grammatical structures:

"No, Missy," said the nurse: "you are to share this young lady's room," designating **me**. Missy did not leave her seat, but I saw her eyes seek **me**.

И она указала на меня. Мисси не встала с места, но отыскала меня глазами.

In the above example, the epiphoric repetition of the pronoun *me* (@#меня#@) in the Russian text loses its final position due to the change of the subject of the subordinate clause (*I saw her eyes seek me* is translated as @#Мисси отыскала меня глазами#@, word-for-word: *Missy found me with her eyes*).

Replacing infinitive verb forms with imperative ones can also prevent preservation of epiphoric repetition in translation:

Go to sleep. I cannot go to sleep. *Спи! — Я не могу спать.*

A more direct translation of this fragment could preserve the rhythmic device: @#Иди спать! — Я не могу пойти спать.#@

Changing parts of speech, as the noun *love* is changed to @#*влюблен, любят, любовь*#@ in the example below, replaces the original epiphora with polyptoton and derivation:

He perhaps was not in **love**; but how many people *Он, по-видимому, не был **влюблен** в нее, но разве так уж много*
ever do **love**, or at least marry for **love**, in this *людей в этом мире **любят** по-настоящему или женятся по*
world. *любви?*

In such cases the rhythmic device is not retained but is changed to a device of a different status.

The text of “The Black Prince” by Iris Murdoch contains 149 epiphora uses, for which 35 correspondent uses are found in the translated text. In 24 contexts, the original epiphora is replaced with other types of repetition in the translated text, most frequently with epanalepsis (13 cases):

You’d better get your own doctor **tomorrow**. Oh, I think I *Вам надо будет **завтра** самому вызвать доктора.*
shall be better **tomorrow**. *– О, **завтра**, я думаю, мне будет лучше.*

The next most frequent replacement is semantic repetition (using a synonym), sometimes accompanied by changes in the word order (5 cases):

Roger has become a **devil**. Some sort of **devil**. *В Роджера просто **дьявол** вселился. Какой-то демон.*

In 2 cases, epiphora is replaced with anadiplosis and gradational repetition. The translation also contains epiphora uses that are not present in the corresponding original sentences, which may be an attempt to partially compensate for the epiphoras that were lost in translation.

Anadiplosis

The four instances of anadiplosis found in Charlotte Brontë’s “Villette” are accurately reproduced in the translated text. It is rather difficult to distinguish repetition at the junction of sentences or parts of sentences from reduplication, which is considered as repetition of the same word in the beginning of a sentence, so in translation one is often replaced with another:

Her features worked, – “**I am sorry; I am sorry!**” ***Прости, прости.***

Besides, the developed application often mixes these rhythmic devices up, which disturbs the statistics. In this situation, the task of separating these devices is on the researcher.

In translations of anadiplosis, the repetition can be reproduced with changes in positions of its components and some changes in the composition of the repeated construction:

I grant I was not looking well, but, on the con- *К этому времени нервы у меня расшатались, но немного, совсем*
trary, thin, haggard, and hollow-eyed; like a *немного. Полагаю, и выглядела я неважно: худая, изможденная,*
sitter-up at night, like an overwrought servant, *с ввалившимися глазами, похожая на сиделку, проводящую ночи у*
or a placeless **person in debt**. **In debt**, however, *постели больного, на переутомленную служанку или **запутавше-***
I was not; (...) ***гося в долгах безработного. Однако я не запуталась в долгах и не***
оказалась в крайней бедности безработного (...)

Anadiplosis can also be replaced with anaphora in translation:

He is fond of **you**. **You** are his favourite. ***Ты** ему нравишься. **Ты** его любимица.*

In this case, neither the number of repeated elements nor their position in the text is changed, but the position of the elements of the device is changed to the beginnings of the (parts of) sentences.

Anadiplosis might be lost in translation:

“**I? I couldn’t do it!**” (*sinking into a chair*). *“**Я? Ни за что в жизни**” (падает в кресло).*

Out of the 37 uses of anadiplosis found in Iris Murdoch's "The Black Prince," two are accurately reproduced in the translated text, for example:

Art is **imagination**. **Imagination** changes, fuses.

Искусство — это воображение! Воображение пресуществляет, плавит в своем горниле.

As in the translation of "Villette," anadiplosis is sometimes replaced with a different device, for example, epanalepsis:

I mean to indicate the curiously controlled rhythmic screaming which goes with a certain kind of **hysterics**. **Hysterics** is terrifying because of its willed and yet not willed quality.

Этим термином я обозначил странно ритмичные, рассчитанные вопли, сопровождающие некоторые истерические состояния. Истерика пугает тем, что она произвольна и произвольна в одно и то же время.

Epanalepsis

Epanalepsis is the rhythmic device the most frequently used in Charlotte Brontë's "Villette." The number of repeated elements is preserved in the translated text, but their position is slightly changed, which affects the rhythm to some extent:

Bretton and I sat alone in the drawing-room waiting her coming; John Graham **Bretton** being absent on a visit to one of his schoolfellows who lived in the country.

*В гостиной остались лишь миссис **Бреттон** и я, так как Джон Грэм **Бреттон** гостил в деревне у своего однокашника.*

The range between the repeated elements is two times smaller in the translated text compared to the original (12 and 6 words respectively), which increases the rhythm of the Russian text.

In the next example, the position of the repeated elements is also slightly changed, but the original version is accompanied by an increase in the number of syllables in the syntagms: 4–5–6, which contributes to the phonetic level perception of the rhythm:

Tie my **sash** straight; make my hair smooth, please. Your **sash** is straight enough.

*Пожалуйста, поправьте мне **пояс** и пригладьте волосы. — Но **пояс** у вас в порядке.*

A comparative analysis of the original and translated contexts of Iris Murdoch's novel "The Black Prince" led to the conclusion that epanalepsis is fully reproduced in the translated text in about 25% of the cases (834 contexts):

I learnt **later** with abhorrence that **he** had set up in business as a self-styled psychoanalyst. **Later** still I heard **he** had taken to drink.

Позже я с отвращением узнал, что он завел себе практику в качестве самозваного "психоаналитика". Еще позже я слышал, что он пьет.

If the number of the repeated elements is relatively large (more than 4), sometimes one or two elements are lost in translation.

In another 5% of the cases (167 contexts), epanalepsis is replaced with another rhythmic device; however, that still preserves the rhythmic pattern of the original text. In some of the analyzed examples epanalepsis was found to be replaced with semantic repetition in combination with syntactic parallelism and positional compensation of epanalepsis, for example:

He was stout (the raincoat failed to button) and not tall, with copious greyish longish frizzy hair and a round face and a slightly hooked nose and big **very** red lips and eyes set **very** close together.

Он был толст (макинтош явно не застегивался), невысок ростом, волосы густые и курчавые, давно не стриженные, с проседью, лицо круглое, со слегка крючковатым носом, толстыми, очень красными губами и удивительно близко посаженными глазами.

Replacement with anadiplosis with positional compensation of epanalepsis:

Real **bears**, I believe, **have eyes** rather wide apart, but caricatured **bears** usually **have** close **eyes**, possibly to indicate bad temper or cunning.

Он походил, как я потом подумал, на карикатурного медведя. Не на настоящего — у настоящих медведей глаза, по-моему, расставлены широко, а вот на карикатурах их рисуют с близко посаженными глазами — вероятно, для того, чтобы выразить их свирепость и коварство.

Polysyndeton

In the translation of Charlotte Brontë's "Villette," the polysyndeton contexts correspond to the originals in 24 cases out of 179, but the exact reproduction of the rhythmic device is found in only 2 of them; the conjunction that ensures the reproduction of this device the best is *and* (@#i#@):

My visits to her resembled the sojourn of Christian and Hopeful beside a certain pleasant stream, with "green trees on each bank, and meadows beautified with lilies all the year round." The charm of variety there was not, nor the excitement of incident; but I liked peace so well, and sought stimulus so little, that when the latter came I almost felt it a disturbance, and wished rather it had still held aloof. One day a letter was received of which the contents evidently caused Mrs. Bretton surprise and some concern.

Жизнь моя не отличалась пленительным разнообразием и волнующими приключениями, но мне нравился этот покой, и, избегая всяческих перемен, я даже любое письмо воспринимала как нарушение привычного хода вещей и предпочитала, чтобы оно вовсе не приходило. Однажды миссис Бреттон получила письмо, содержание которого явно удивило и несколько обеспокоило ее.

Polysyndeton is quite successfully reproduced in the translation of Iris Murdoch's novel "The Black Prince": it is reproduced accurately in about 45% of the contexts (98 contexts). It should be noted that additional polysyndeton uses were found in almost all of the translated contexts: 219 examples in the original text and 366 in the translated text, for example:

Both taxman and dentist only too readily image forth the deeper horrors of human life: **that** we must pay, perhaps ruinously, for our pleasures, **that** our resources are lent, not given, and **that** our most irreplaceable faculties decay even as they grow.

И зубной врач, и налоговый инспектор, естественно, символизируют для нас подспудные ужасы жизни; они говорят о том, что мы должны платить, даже если цена разорительная, за все наши удовольствия, что блага даются нам в долг, а не даруются, что наши самые невосполнимые богатства гниют уже в процессе роста.

The tool identifies any occurrences of identical conjunctions within the same paragraph as polysyndeton, which somewhat expands the definition of the rhythmic device, which is now not limited to repetition of conjunctions between coordinate elements, but can encompass several groups of coordinate elements.

Symploce, epistrophe, and reduplication

No uses of symploce, epistrophe, or reduplication were found in Charlotte Brontë's novel "Villette."

One of the nine uses of symploce found in Iris Murdoch's "The Black Prince" has a corresponding reproduction in the translated text:

I turned back and drove the other way, through the village, past **the church**. I even stopped and went into **the church**.

Я повернул назад и поехал через деревню мимо церкви. Я даже остановился и зашел в церковь.

The other occurrences of the rhythmic device were not lost, but underwent stylistic transformations. Thus, in five of the eight remaining cases, symploce is replaced with epanalepsis:

She kept drawing it out without telling me and **buying clothes**. **She** went mad over **buying clothes**.

А она втихомолку тянула фунт за фунтом и покупала себе тряпки. Она на этих тряпках просто помешалась.

In two cases symploce is replaced with a combination of epiphora and epanalepsis:

What did you say? What could I say? *А вы что ответили? Что я могла ответить?*

In one case, it is preserved as syntactic parallelism:

I write whether I feel like it or not. I complete things whether I think they're perfect or not. *Я пишу независимо от того, легко мне или трудно. И завершаю любую работу, удалась она мне или не удалась.*

Three independent uses of symploce were found in the translated text, possibly employed by the translator as compensation for the lost rhythmic devices.

The original text contains 23 uses of reduplication, two of which have corresponding versions in the translated text. The tool has not found any uses of epistrophe in the translated text.

RESULTS AND DISCUSSION

It should be noted that this detailed analysis of the large number of rhythmic aspects (almost two thousand aspects in one of the texts and more than four thousand in the other) only became possible with the use of search automation and visualization of results. The developed software tool allowed linguists to focus on the analytical part of the study, compare each aspect of text rhythmicity between the original and translated texts in more detail, and draw conclusions about the reasons for losses or replacements of these aspects.

Quantitative analysis of the reproduction of rhythmic devices in the translation of Charlotte Brontë's novel "Villette" showed that such reproduction was the closest to the original in the case of anadiplosis (33.3%), followed by epiphora (20%), anaphora (14.7%), epanalepsis (7.2%), and polysyndeton (1.1%). In general, the percentage of correspondences in the use of rhythmic devices is quite low—7.1%. In the translation of Iris Murdoch's "The Black Prince," the completeness of reproduction of rhythmic devices is different: polysyndeton (44.7%), simploce (33.3%), epanalepsis (30%), epiphora (23.5%), anaphora (17.4%), reduplication (8.7%), and anadiplosis (5.4%). The overall percentage of correspondences is significantly higher: 28.2%.

Of all the considered devices, anaphora is the most accurately reproduced in translation in terms of position in the text and the number of repeated elements. The large number of aspects of this rhythmic device found in the texts made it possible to generalize and formulate the following reasons for anaphora losses.

1. Impossibility of reproduction due to the objective discrepancy between grammatical capabilities of English and Russian. In the analyzed examples, that often relates to differences in the use of verb voice: for example, in the Russian language passive voice cannot be used with intransitive verbs and verbs with prepositional government, while, in the English language, such use is normal:

I was upset and annoyed when my father once approached the subject, and although I could see that he had been made utterly miserable, I resolutely refused to discuss it. *Когда мой отец попытался однажды заговорить со мной о Присцилле, мне это было крайне неприятно, и я, хотя и видел, как он расстроен, решительно отказался обсуждать эту тему.*

In this example one of the elements of the anaphoric repetition of *I* is lost in translation.

2. Risk of violating the usage (the use of language units conventionally accepted by native speakers of a given language) of the target language. For example, that would be the motivation behind translating the phrase *He needed* as @#ему был нужен#@ (word-for-word: *for him was needed*) and not directly as @#он нуждался#@.

3. The presence of formulaic expressions (clichés):

I am 58. I still dream about it at least once a week. *Мне 58. Он и по сию пору снится мне примерно раз в неделю.*

4. "Aggressive" (obtrusive) repetition of 5–6 elements can be a style violation. In addition, repetition of subject pronouns (used in English texts in abundance, especially *I*) is redundant for the Russian lan-

guage, where the relevant grammatical categories of person and number are additionally expressed through verb conjugation, changes in root vowels and consonants, etc.

I waited, I listened.

Я выждал, вслушивался.

Loss of contexts with epiphoric repetition is expected, due to the differences in the focus (the part of the sentence that contains new information about the topic) position in English and Russian. For example, in English, personal pronouns can be unstressed and semantically unimportant when used at the end of a sentence or phrase (where they can act as epiphora elements), while in Russian the final word of a sentence or phrase is automatically assigned phrasal stress and perceived as the focus. To preserve the correct topic–focus division of a sentence or phrase, the translator often has to rearrange the elements of that sentence or phrase, which inevitably leads to epiphora losses. In a significant number of contexts, that is compensated with the use of other rhythmic devices.

Anadiplosis is in most cases lost without any compensation in the rhythmic structure of sentences. However, the total number of anadiploses in the translated texts suggests that this rhythmic device is used equally often in the original and translated texts, albeit with positional compensation: in the case of Charlotte Brontë's "Villette," the original text contains 12 uses of anadiplosis and the translated text contains 4 that correspond to the original and another 23 that were used by the translator independently; in the case of Iris Murdoch's novel, for the original 37 uses there are 4 that fully or partially correspond to the original and 33 that were used by the translator independently.

Epanalepsis is the most frequent rhythmic device in the analyzed texts. In most cases epanalepsis is lost in translation, which is then compensated with its use in sentences that do contain epanalepses in the original. Most often, these original sentences contain other rhythmic devices, in several cases they do not contain any repetition at all.

One of the most frequent causes of loss of epanalepsis is the fact the use of the personal pronoun *I* as a subject, as well as possessive pronouns, especially in preposition to nouns denoting body parts, clothing, kinship terms, etc., cannot be reproduced consistently in all cases, as that would violate usage.

Polysyndeton is the most consistently reproduced in the translation of Iris Murdoch's novel "The Black Prince." In many contexts, this device is compensated with semantic repetition (using a synonymous conjunction) or is lost, for example:

The shop sold daily papers **and** magazines, writing paper **and** so on, **and** horrible gifts. Они продавали газеты **и** журналы, бумагу всевозможных сортов, **а также** безобразные "подарки".

The other rhythmic devices occur in the original texts quite rarely or not at all, which precludes significant conclusions about characteristics of the translation of these aspects. It is of interest to note that although reduplication is completely absent in the original text of Charlotte Brontë's novel "Villette," a rather large number of its uses are found in the novel's translation. That might be a feature of the translator's individual style.

In general, the main reason for the low completeness of reproduction of rhythmic devices in English to Russian translation is the specifics of the lexical and grammatical structures of the source and target languages, which belong to different language groups. Additional factors that contribute to losses or replacements of aspects of text rhythmicity are, for example, strategies chosen by translators, as well as quality of translation and adequacy of reproduction of rhythmic characteristics. A detailed analysis of these factors requires further research.

It should be noted that research and experiments of this kind are highly labor-intensive. It is especially difficult for a human researcher to perform both a laborious search for a large number of various rhythmic devices that occur in texts with drastically different frequencies and a comparative analysis of the original and translated texts. The developed software tool ProseRhythmDetector demonstrated high efficiency in performing the time-consuming task of searching for the rhythmic devices anaphora, epiphora, symplote, anadiplosis, epistrophe, reduplication, epanalepsis, aposiopesis, and polysyndeton. Crucially, it also provides a visualization of search results, so the expert can easily navigate the analyzed texts. The tool made it possible to conduct an experiment on large literary texts, which would be almost impossible manually, and assess the rhythmicity of the analyzed texts in terms of lexical and grammatical levels of rhythm implementation.

CONCLUSIONS

In this study, the authors constructed algorithms for such complex linguistic concepts as devices of prose rhythmicity. The developed software tool automated a significant part of natural language text processing tasks related to the search and enumeration of an entire complex of rhythmic aspects of literary texts. Assisted by the tool, researchers conducted a comparative analysis of the original and translated texts of two voluminous novels and drew conclusions about the characteristics of the reproduction of prose rhythmicity in English to Russian translation.

The degree of correspondence or discrepancy between rhythmic devices in the original text and their reproductions in the translation is a marker of translation adequacy, but mostly it highlights the specifics of language structures in general, especially those that concern the nature of reproduction of the considered rhythmic devices, their forms, variations, substitutions. In this way, rhythm can be evaluated as a complex phenomenon and possible ways of translating certain rhythmic devices can be generalized.

FUNDING

The study was supported by the Russian Foundation for Basic Research, project no. 19-07-00243.

REFERENCES

1. Golyakova, L.A., The rhythm of a literary work of art: A communicative-pragmatic aspect, *Vestn. Permsk. Univ., Ross. Zarub. Filol.*, 2011, no. 3, pp. 94–99.
2. Golubeva-Monatkina, N.I., On the problem of prose rhythm, *Izv. Ross. Akad. Nauk, Ser. Liter. Yazyka*, 2017, vol. 76, no. 2, pp. 16–27.
3. Golubeva-Monatkina, N.I., On the rhythm of literary prose in the light of M.M. Bakhtin's ideas and translation, *Russ. Yazyk Kul't. Zerkale Perevoda*, 2016, vol. 1, pp. 105–113.
4. Ivanova-Luk'yanova, G.N., The rhythm of prose in reality and perception, *Tr. Inst. Russ. Yazyka im. V. V. Vinogradova*, 2016, vol. 7, no. 7, pp. 365–382.
5. Fenenko, N.A., Linguocultural adaptation of the text in translation: Limits of the possible and the permissible, *Vestn. Voronezh. Gos. Univ., Ser.: Lingvist. Mezhdkul't. Kommun.*, 2001, vol. 1, no. 7, pp. 70–75.
6. Delmonte, R., Computing poetry style, *Proceedings of 1st International Workshop ESSEM 2013/CEUR Workshop Proc.*, 2013, pp. 148–155.
7. Bobenhausen, K. and Hammerich, B., Literary metrics, linguistic metrics, and the algorithmic analysis of German poetry using Metricalizer (2), *Langages*, 2015, vol. 3, pp. 67–88.
8. Niculescu, I. and Trausan-Matu, S., Rhythm analysis of texts using Natural Language Processing, *RoCHI*, 2016, pp. 107–112.
9. Niculescu, I. and Trausan-Matu, S., Rhythm analysis in chats using Natural Language Processing, *RoCHI*, 2017, pp. 69–74.
10. Kishalova, L.V., Analysis of the features of the rhythmic structure of texts of different styles of speech, *Vestn. Bryansk. Gos. Univ.*, 2016, vol. 27, no. 1, pp. 257–261.
11. Boychuk, E., et al., Automated approach for rhythm analysis of French literary texts, *Proceedings of 15th Conference of Open Innovations Association FRUCT*, 2014, pp. 15–23.
12. Damasevicius, R., Kapociute-Dzikiene, J., and Wozniak, M., Towards rhythmicity analysis of text using empirical mode decomposition, *Proceeding of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2017, vol. 1, pp. 310–317.
13. Kelly, A.R., et al., Toward an ontology of rhetorical figures, *Proceedings of the 28th ACM International Conference on Design of Communication*, 2010, pp. 123–130.
14. Hromada, D.D., Initial experiments with multilingual extraction of rhetoric figures by means of PERL-compatible regular expressions, *Proceedings of the Second Student Research Workshop associated with RANLP*, 2011, pp. 85–90.
15. Toldova, S., et al., Error analysis for anaphora resolution in Russian: New challenging issues for anaphora resolution task in a morphologically rich language, *Proceedings of the Workshop on Coreference Resolution beyond OntoNotes*, 2016, pp. 74–83.
16. Dubremetz, M. and Nivre, J., *Rhetorical figure detection: Chiasmus, epanaphora, epiphora*, *Front. Digital Humanit.*, 2018, vol. 5.

Translated by A. Ovchinnikova

SPELL: 1. OK