Automated Approach to Rhythm Figures Search in English Text

Elena Boychuk[1], Inna Vorontsova[1], Elena Shliakhtina[1], Ksenia Lagutina[2], Olga Belyaeva[1]

[1] Yaroslavl State Pedagogical University named after K.D.Ushinsky
Respublikanskaya Str. 108/1, 150000, Yaroslavl, Russia
[2] P.G. Demidov Yaroslavl State University,
Sovetskaya Str. 14, 150003, Yaroslavl, Russia
elena-boychouk@rambler.ru, arinna1@yandex.ru, ElenaV_Yar@mail.ru,
lagutinakv@mail.ru, olbelyaeva@yandex.ru

Abstract. Text rhythm is recognized as being one of the most important subject areas of modern linguistic studies. There is a considerable amount of literature on the analysis of rhythm in poetry and literary prose. However, few researchers have addressed the problem of using automated tools for rhythm analysis, whereas automated methods can be of great benefit to this cause, especially when the research is conducted on large text corpora. This paper presents a new automated approach to integrated search of rhythm figures in fiction including anaphora, epiphora, anadiplosis, symploce and simple repetition provided for by an original lexical tool designed within the framework of the research. The ad hoc experiments have proved this approach to be reliable and informative.

Keywords: text rhythm, rhythm analysis, natural language processing, rhythm figures, automated approach to rhythm analysis

1 Introduction

In terms of linguistics rhythm is a literary device that demonstrates the long and short patterns through stressed and unstressed syllables, particularly in verse form [1]. However, rhythm has a more complex structure that can manifest itself at various linguistic levels in various text types and is characterized by a special "movement" (mouvant) [13]. Fictional text rhythm has so far lacked a clear definition that can be followed without reservation. Nonetheless in this paper we define rhythm as regular repetition of similar and commensurable units of speech that performs structuring, text-forming, and expressive functions [6].

The main purpose of rhythm analysis is deep penetration into the creative method of an author, into their intent, originality of individual creativity and skill. Identifying the specificity of the rhythm of writer's works makes it possible to more successfully solve the problem of determining the authorship of texts. This method is widely used in poetic text analysis, while its application to the research of fiction can be questionable [9]. The problem confronted is large text

processing. Therefore, development of automated tools for rhythm analysis in a non-poetic text is among the primary tasks of computational linguistics.

2 Problem statement

The phonetic aspect of rhythm analysis is usually defined as the sonic periodicity, i.e. change of consonants and vowels and reiteration of the same sounds. It also includes the analysis of accent distribution, pauses and tempos. The lexical aspect of rhythm involves repetition of words, for example, at the beginning or the end of sequences. Within the framework of a grammar aspect of literary stylistics, morphological and syntactic means of rhythmization are distinguished for the analysis of rhythm. For this analysis, such characteristics as the repeated use of words with a common root, the repetition of words in different tenses, rhetorical questions, exclamations, homogeneous sentence members are important.

The paper considers an original software application designed for the automated search for lexical figures of repetition. The application is set to be simple and convenient focusing on finding the figures of repetition and displaying them on the screen.

The following figures of repetition are selected for the research:

- anaphora (repetition of words at the beginning of a sentence or clause);
- epiphora (repetition of words at the end of a sentence or clause);
- symploce (joint use of anaphora and epiphora);
- anadiplosis (repetition of the final word of one clause or sentence at the beginning of the following clause or sentence);
- simple repetition.

Summarily, the application should perform the tasks set above and be easy to use.

3 Existing tools: state-of-the-art

Most of works in the field of rhythm analysis look at the definition of phonetic aspects. American researchers Green et al. [11] analyze the rhythm of a poetic text. They define the analysis as extraction of patterns from existing online poetry corpora. They use these patterns to generate new verses and translate the existing poems.

The "Ritminme" program (`http://www.ritminme.ru`) allows to evaluate the poetic rhythm selecting a rhyme to the given word and defining a rhythm scheme basing on revealing the stressed and unstressed syllables.

Kishalova [12] proposes a similar approach. She analyzes the rhythm in Russian texts with the "PULSE" program. The tool considers the average factor of unstressed syllables when analyzing the rhythmic structure of the text. Experiments with fiction, news, and scientific texts show the dependency of the rhythm on the text style.

Methods adopted in syntactical analysis are also applicable to rhythm analysis of poetry and prose.

Belousov and Dusakova [4] propose the tool that helps to analyse text rhythm basing on the automatic computation of sentence lengths.

Couranjou and Lachambre [7] analyze the grammar structure of sentences and their rhythm. The syllables are counted and a rhythm curve is built with regards to the syllable number in a text portion, the research is based on a case study of fiction.

Toldova et al. [17] compare systems of anaphora detection. The best results are shown by linguistic algorithms that apply approaches based on rules and ontologies.

Dubremetz and Nivre [8] use a binary logistic regression classifier to extract chiasmuses, anaphoras, and epiphoras from political texts. This system proves to be quite efficient: for epanaphora the accuracy and F-measure are around 55–63 %, for epiphora — around 50 %, for chiasmus — 78.3 %.

The best approaches to text rhythm analysis enable experts to make their decisions on the basis of several types of text or word features.

In the area of rhythm analysis in Russian, English and French, the "Rhymes" program (`http://rifmovnik.ru`) by N. Ketsaris assists in text processing in order to find rhymes for the specified word by phonetical and lexical characteristics.

Russian researchers Boychuk et al. [6] analyze the rhythm of French fiction. Their program allows to do it by applying a number of methods grouped according to the aspect of analysis: phonetic, lexical and grammar.

Balint et al. [3,2] propose a method of English text rhythm evaluation. It analyzes different rhythmic features, including organizational, lexical, grammar, phonetic, and metrical. The program that implements this method uses statistical algorithms that allow to achieve around 80 % accuracy in prediction of the text genre by its rhythm features.

Niculescu and Trausan-Matu [14,15] describe a Natural Language Processing application that analyzes the rhythm of English, Romanian, and French texts of different styles: poems, fiction, and political speech. It automatically performs word hyphenation, search of stressed and unstressed syllables using a dictionary, and detection of phonetic, metrical, and grammar rhythm features. The authors show the benefits of their application for the automatization of linguistic research.

The existing instruments prove to be targeted at the analysis of text rhythm at the phonetic and lexical levels and/or at the assessment of sentence length. The novelty of the tool described in this paper is seen in its ability to search and process stylistic devices based on repetition.

4 The approach

The basis of the rhythmic means of this tool is a repetition that has the following structure: source element + repeating elements. Depending on the means, the

elements of repetition can have the initial, final, middle, contact, non-contact, cross positions within a rhythmic unit.

In order to achieve a higher degree of rhythmization with the help of these figures, we should set out the following conditions: 1) presence of several repeating elements, 2) close proximity of the original and repeating elements, 3) high frequency of elements in the text.

To analyze the use of figures of rhythmization, it is necessary to clarify some stylistic terms since they allow for different interpretations in dictionaries and academic literature.

Anaphora is one of the most complex figures. Firstly, it does not have a clear distinction from the concept of deixis, often represented as a category combining anaphora and deixis itself. In this combination, anaphoric relationships indicate the context elements when referring from one word or phrase to another. This relationship is called an associative anaphora but as for rhythm analysis a full (tautological) anaphora is most important.

Secondly, in stylistics there are several types of anaphora as a rhetoric figure depending on the ways of its manifestation: phonetic, lexical and syntactic. At the same time, some dictionaries consider a syntactic anaphora as a lexical subtype, and others as an independent type. In this paper, a lexical anaphora is regarded as most indicative from the rhythm viewpoint. Two elements of anaphora repeated in adjoining sentences are sufficient for text rhythm perception.

Anadiplosis and simple repetition appear to be most confusing in terms of distinction.

Most linguists share an opinion that anadiplosis (Greek "ana" — again and "diploos" — double) is a repetition of the final word of one clause at the beginning of the following clause or the repetition of the final word in a sentence and the beginning of the next sentence [10]. However, a different opinion exists. For example, French linguists D. Bergez, J.-J. Robrieux say: "Cette figure de répétition consiste à reprendre dans une phrase (souvent au début) un mot ou un groupe de mots de la phrase précédente, de maniére à établir une liaison" (This stylistic figure consists in repeating a word or a group of words of one sentence at the beginning of the next sentence) [16,5]. Thus, in this definition, the number of elements within anadiplosis is not determined.

The tool discussed in this paper also offers the implementation of a simple retry search. At this level of development, such types of repetitions as reduplication and epanalepsis are not distinguished.

The latter is one of the most controversial in stylistics, since, depending on the conditions of use, it can be identified with anaphora, epiphora, anadiplosis and other stylistic figures. The term "epanalepsis" is used as synonymous with simple repetition, while there is a definition of etymological paraphrase for both figures [10]. The main stylistic function of these figures is the expression of an emotional state, certain feelings: anger, pain, despair, joy, etc.

In this paper we consider epanalepsis as a repetition of a word or a part of a statement interspersed with intermediate words.

Besides, it is important to determine the number of the elements. The smaller the distance between repetitions, the more rhythmic the text is. Epanalepsis is defined as a figure that excludes contact repetitions at the junction of clauses or at the junction of sentences (anadiplosis functions), as well as cases of non-contact repetitions at the beginning (anaphora) or the end (epiphora) of clauses and sentences, and simultaneously at the beginning and at the end of clauses (symploce).

It is also necessary to clearly distinguish between epanalepsis and reduplication (palilogy). Both are based on repetition: the repetition of words in the contact position to accentuate their semantics is called reduplication, while the repetition of words and phrases after intermediate words is considered as epanalepsis.

The algorithm for analyzing the entered text when identifying a lexical anaphora, epiphora, or symploce includes the search for repetitions at the beginning, the end of sentences, as well as at the junction of sentences and clauses. To locate the position of the desired word, a punctuation criterion is applied (comma, period, semicolon, exclamation and question marks, three dots).

To identify anadiplosis, the tool searches for repeating words separated by commas without changing the forms, regardless of the position in a sentence. To reveal the reduplication, the tool searches for a repetition of words separated by commas without changing word forms at the beginning of a sentence. When working with epanalepsis, it is necessary to take into account the repetition of words that are not in a contact position in relation to each other, but are located within a small rhythmic unit.

5 Software implementation of the tool

A randomly selected text is uploaded and automated search for the five figures described above (anaphora, epiphora, anadiplosis, symploce and simple repetition) starts. The elements are highlighted with the corresponding colors, and a list of all words in the text and the number of their repetitions are displayed.

As a result, the researcher can receive information about the used lexical figures in the text, propose a theory about the text authorship, evaluate the quality of translation and analyze the rhythm of the text.

The functional tool has a simple interface for displaying the text and the list of all its words with the number of repetitions. Also it finds specific lexical figures in the text, highlights the found elements and display their list separately 1.

The program was implemented on the JavaScript language using HTML and CSS. It is available at `https://github.com/text-processing/html-tool`.

The word list with the number of repetitions is formed according to the following algorithm. In the beginning, all non-letter characters characters are replaced with spaces. Extra spaces are removed, i.e. where their number is more than one. Then the whole text is divided into an array, where the elements are individual words. In the resulting array the algorithm counts the number of repetitions of each element. As a result, we get an associative array of words associated with the number of their repetitions. The elements of the resulting array

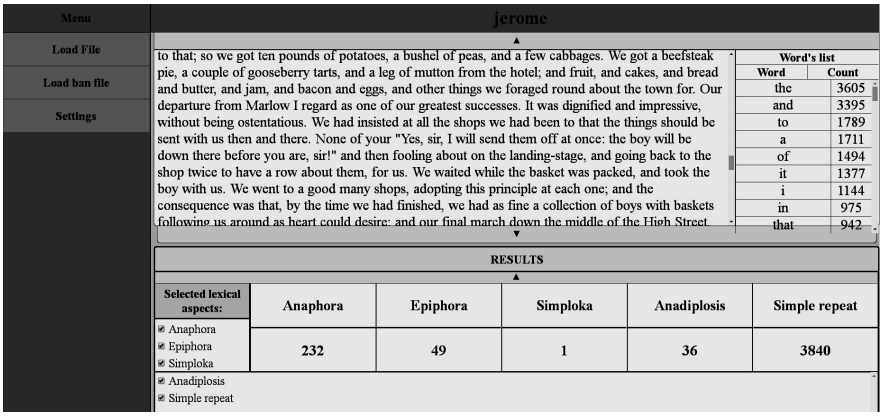

Fig. 1. Main page of the application

are sorted in descending order by the number of repetitions and are displayed in a table.

The user can edit the list of displayed words, i.e. delete specific words that are not needed for investigation. The list is formed according to the algorithm described above, but during the element-by-element output of the result, the displayed item is searched for in the stop list (the list of stop words, i.e. ignored words). If the element is found, it is not displayed.

The tool also allows to search lexical figures.

The search for anaphora in the text occurs according to the following algorithm. In the beginning, it forms an array of the first words of each sentence in the text. This array counts word repetition number. If it exceeds one, the word is entered into the resulting array. Then a new array is created from phrases the beginning of which is formed of words from the resulting array. In this array we search for repeating elements again. If repetition number is less than two, the element is deleted. The remaining elements form the resulting array by updating the existing elements or adding new ones. In the end, we get a list of anaphoras.

The algorithm for searching epiphoras is similar. The significant difference is that the array is formed from the last words of the sequences. At the same time, the formed phrases contain words from the end of this array.

After extraction of these two figures, the application searches symploce choosing cases when anaphoras and epiphoras appear together.

The following algorithm is applied to the search of the anadiplosis. The program searches elements by the pattern: “word / phrase + punctuation mark + word / phrase” and adds them into the resulting array. Then the algorithm compares the left and right parts separated by a punctuation mark. If they are identical, the element is added to the resulting array.

The search for simple repetition is carried out according to the following algorithm. In the beginning, all non-literal characters are replaced with spaces. Then the whole text is divided into an array of individual words. For every word

Table 1. Results of lexical aspects search

Means of rhythmizaton	Number of occurrences in texts	Accuracy
Anaphora	4407	89
Epiphora	2564	93
Symploce	18	65
Anadiplosis	458	62
Simple repetitions	28542	93

the algorithm counts the number of repetitions and deletes words when this number is less than two. In the end, we get an array of simple repetitions.

6 Experiments

The experiment was conducted by a research team based at the Department of Foreign Languages, Yaroslavl State Pedagogical University named after K.D. Ushinsky. The text corpus was derived from the works of 23 English authors selected with regards to the manifestation of such rhythmic figures as anaphora, epiphora, anadiplosis, symploce and simple repetition. The following writers were chosen for that purpose: K. Atkinson, J. Austen, Ch. Bronte, Ch. Dickens, E.M. Foster, J. Fowels, N. Gaiman, E. Gaskell, Th. Hardy, J. Joyce, D.H. Lawrence, D. Lessing, D. du Maurier, I. McEwan, I. Murdoch, S. Muriel, T. Pratchett, J.K. Rowling, R.L. Stevenson, S. Thomas, J.R.R. Tolkien, O. Wilde, V. Woolf.

The researchers were divided into 2 groups, the first of which processed the text manually, the second - used the Rhythmanalysis tool. The main objectives of the experiment were 1) to determine the efficiency of the tool in terms of time and accuracy of results, 2) to detect errors when working with the tool, 3) to analyze problems and devise ways to solve them.

The first group of 8 researchers worked for 32 days 2 hours a day. The second group (also 8 people) coped with the task within 4 days, working 2 hours a day. Thus, the efficiency of the application is obvious: the text processing time is reduced by 87.5 %, which allows the researcher to be spared from a lot of tedious and monotonous work.

The second stage of the experiment implied the identification of errors, which the tool allows when searching for rhythmic figures (anaphora, epiphora, symploce, anadiplosis, simple repetition) in the texts. At this point, the re-searchers had to spend more time checking on the data provided by the tool, since they had to work both with the tool and with the texts.

The results of the experiments are presented in Table 1. In the analyzed texts the tool found 4407 anaphoras. 89 % out of them were detected correctly. For example, *"**I wanted** a miracle job advertisement. **I wanted** someone to come along and say, "Just do what you're good at and we'll give you enough money for your rent, bills, cigarettes and some nice food and clothes"* (Scarlett Thomas).

In the course of the analysis we also found a few disputable cases of anaphora. First of all, not all anaphora elements were detected by the tool. In the following

case it is caused by the word number limit in anaphora: *"**It asks** whether it is possible or even desirable to disrupt this. **It asks** whether it is possible to find meaning in a world overflowing with it"* (Scarlett Thomas). Another reason for the partial detection of anaphora is the presence of an extra element within the analyzed unit:

*"**Not in** Sylvie's **room** (they had long ago ceased to think of it as a room that belonged to two parents).*

***Not in** Maurice's **room**, so generously sized for someone who spent more than half his life living at school."* (Kate Atkinson)

However, the displayed text fragment allows to manually detect all elements of the anaphora.

Not all displayed cases of the pronoun **he** can be regarded as anaphora. For example, *"**He** was a member of a cycling club and every Sunday tried to wheel as far away from Birmingham's smogs as he could, and he took his annual holiday by the sea so that he could breathe hospitable air and think himself an artist for a week.*

***He** thought he might try to put some figures in his painting, it would give it a bit of life and 'movement', something his night-school teacher (he took an art class) had encouraged him to introduce into his work"* (Kate Atkinson).

There is a text fragment of considerable length between the pronouns. Moreover, the first pronoun is in the middle of the paragraph while the second one is at the beginning of a new paragraph.

In the following example the displayed preposition cannot be reckoned as anaphora because the word is used in different meanings: *"**At** first, being little accustomed to learn by heart, the lessons appeared to me both long and difficult; the frequent change from task to task, too, bewildered me... **At** that hour most of the others were sewing likewise; but one class still stood round Miss Scatcherd's chair reading..."* (Charlotte Bronte).

In 29 cases simple repetition was displayed as anaphora. For example:

- *Eight years.*

- *Eight years!* (Charlotte Bronte)

In 21 % of cases the detected units were not anaphoras. Here, it is important to emphasize that all cases of the definite article usage cannot be referred to as anaphora. For example, *"**The** Librarian jumped it. **The** Luggage, of course, followed them with a noise like someone tapdancing over a bag of crisps"* (Terry Pratchett). Besides, according to the rules of the English language the stress is usually put on meaningful parts of speech. Therefore, the article cannot influence the rhythm of the text in general. The only possible exception is the situation when the article is a part of another rhythm figure, for example, gradation.

The contextual analysis allowed to detect the following number of end-of-a-sequence repetitions provoking a rhetorical effect, for example: 35/46 (76 %) in I. Murdoch's "The Bell", 174/187 (93 %) in D. Du Maurier's "Rebecca" and 70/74 (94.5 %) in V. Wolfe's "To the Lighthouse". We view this statistics to speak for a high sensitivity of the method applied to the search of **epiphora**:

*"Frank **knew**. And Maxim did not know that he **knew**"* (Daphna Du Maurier "Rebecca").

*"The young man was abusing the government. William Bankes, thinking what a relief it was to catch on to something of this sort when private life was disagreeable, heard him say something about "one of the most scandalous acts of the present government." Lily was **listening**; Mrs Ramsay was **listening**; they were all **listening**"* (Virginia Wolfe "To the Lighthouse").

While analyzing the use of **anadiplosis**, the following was revealed:

1. The average number of uses per book is 18-37 units. A total of 458 cases of anadiplosis were investigated. Of these, 62 % are the correct search.
2. Most cases of use are classified by the tool as reduplication (repetition of the same words in the contact position in one clause, whereas anadiplosis is used at the junction of the clauses), for example: *"A **very, very** brief time, and you will dismiss the recollection of it, gladly, as an unprofitable dream, from which it happened well that you awoke. My **little, little** child.' cried Bob. The father of a **long, long** line of brilliant laughs."* (Ch. Dickens)
3. The tool considers repetitions that are characterized by a polysyndeton (the use of coordinating conjunctions close together, and more than needed, for stylistic effect) as an anadloplosis, for example: *"Scrooge went to bed again, **and thought, and thought, and thought** it over and over, and could make nothing of it."* (Ch. Dickens). This case must be regarded as a reduplication with a polysyndeton.
4. The most indicative cases of anadiplosis revealed by the instrument are the cases of using it at the junction of clauses: *"It was right to do **it, it** was kind to do **it, it** was benevolent to do it, and he would do it again. It was right to do **it, it** was kind to do **it, it** was benevolent to do it, and he would do it again."* (I. Murdoch). In this case, the example is also combined with the epiphora (to do it at the end of sentences).

While carrying out the experiment the lexical tool identified the following phrases as **symploce**: *"O! O!", "Norbert! Norbert!", "Why? Why?", "And so? And I'm a hundred times more vulnerable than I was. And so?", "Nothing! Nothing!", "No! No!", "The Lighthouse! The Lighthouse!"*. After a thorough analysis, these examples are considered to be just a simple repetition of all the elements of the phrase which can be seen as reduplication, but not symploce.

As for **simple repetition**, this tool has the highest percentage of manifestations in the texts (more than 1000 cases per work). This is primarily due to the fact that simple repetition as a complex means of rhythm involving the division into different types of repetitions, combines the cases of use of reduplication, epanalepsis, and in some cases anadloplosis. It is possible to layer of one means to another.

7 Discussion

To improve **anaphora** detection it is necessary to

– exclude articles from the list of the checked words;
– reduce the number of the words in a search unit to 20;
– define the anaphora parameters more clearly.

For that purpose it is important to limit the number of words to 2-3 during the search and to introduce the punctuation parameter that will help to distinguish anaphora from simple repetition.

The main reason for misidentifying the **epiphora** is a considerable length of sequences (mainly sentences) where new clauses hamper the perception of the stylistically marked epiphora:

*"... and he could not help hoping that Toby would sooner or later force such a tête-à-tête upon **him**. He wished that somehow he could pull out of this mess the atom of good which was in it, crystallizing out his harmless goodwill for Toby, Toby's for **him**".* (Iris Murdoch "The Bell").

*"He wanted something else that I could not give him, something he had had **before**. I thought of the youthful almost hysterical excitement and conceit with which I had gone into this marriage, imagining I would bring happiness to Maxim, who had known much greater happiness **before**".* (Daphna Du Maurier "Rebecca").

In a few cases (not exceeding 3 in the sample) the repeated words, usually personal pronouns, have different referents, which enables us to consider such repetition as accidental and therefore irrelevant:

*"Her face was glowing and she put up one hand to hide **it**. Her cigarette fell on the floor and she abandoned **it**".* (Iris Murdoch "The Bell")

Other examples (10 in the sample) reveal the inability of the method to detect repetition at the end of clauses along with its successful identification of the epiphora in sequential sentences:

*"Truthfulness **is enjoined**, the relief of suffering **is enjoined**, adultery **is forbidden**, sodomy **is forbidden**.*

*And I feel that we ought to think quite simply of these matters, thus: truth is not glorious, it is just **enjoined**; sodomy is not disgusting, it is just **forbidden**".* (Iris Murdoch "The Bell").

The solution to the problem is expected to be found in designing a feature that would make the tool capable of simultaneously analyzing a sentence-and a clause-long context with the "comma" serving the marker.

As a way of resolving the problem of differentiating between **anadiplosis** and **reduplication** by the tool, three possible ways are proposed:

– formulation of rules for the search for reduplication as a repetition at the beginning of a sequence in steps of 3-4 words
– exclusion of all adjectives (including a comparative degree) and adverbs from the list of anadiplosis;
– restriction of cases of anadiplosis only to those that are observed at the junction of clauses, separated by semicolon, period, question or exclamation marks, as well as three dots.

The main feature distinguishing between the anadiplosis and reduplication with a polysyndeton should be considered as a union which is not an element of anadiplosis. In order to clearly identify structures with a polysyndeton, it seems appropriate to compile a list of all possible unions used at the junction of clauses.

Due to the fact that cases with the pronoun "it" have 67 % of the total number of anadiplosis use, it is proposed to add cases with it to the list of search conditions of the tool when it is considered as a repetition at the junction of clauses separated by commas (it, it).

Taking into account the fact that the lexical tool identifies simple phrases and utterances as **symploce**, but not the repetition of the beginning and the end of the sentence which consists of the subject and/or the predicate, it is considered necessary to apply a restriction (of at least 5 words which should be in a clause/sentence) to the tool while detecting symploce in such texts as fiction.

As a solution to the problem of inaccurate detection of the types of **simple repetitions**, it is necessary to divide these rhythmic figures into reduplication and epanalepsis. As for reduplication, the parameter should be the contact position of words separated by commas (but not at the junction of clauses or sentences), and, as for epanalepsis, the parameter should be the use of identical words at certain intervals.

8 Conclusion

The described tool for the analysis of fiction text rhythm allows to conduct integrated research into lexical rhythm figures including anaphora, epiphora, anadiplosis, symploce and simple repetition. The tool has shown 80.4 % reliability which we view as relatively high. Pitfalls in the process of lexical figures search and evaluation are in a few cases explained by the impossibility of excluding certain conditions of the repetition use. If this occurs, the decision whether the case should be attributed to a lexical figure is taken by the researcher. This automated approach holds the lowest number of pitfalls when applied to the search of anaphora, epiphora and simple repetition. To make the detection of the anadiplosis more precise, simple repetition must be divided into reduplication and epanalepsis, which will spare cross search between the anadiplosis and the simple repetition. In terms of symploce, this repetitive figure is rare in English texts, the search results can be improved by expanding the word range between the repeated clauses. Future work will concentrate on the definition of the author's individual style by means of rhythm figures as well as on the application of the above principles of rhythm analysis to texts written in languages other than English.

Acknowledgements

The reported study was funded by RFBR according to the research project №19-07-00243.

References

1. Literary Devices. Definition and Examples of Literary Terms. Metaphor, `http://www.literarydevices.net/metaphor/`
2. Balint, M., Dascalu, M., Trausan-Matu, S.: Classifying written texts through rhythmic features. In: International Conference on Artificial Intelligence: Methodology, Systems, and Applications. pp. 121–129. Springer (2016)
3. Balint, M., Trausan-Matu, S.: A critical comparison of rhythm in music and natural language. Annals of the Academy of Romanian Scientists, Series on Science and Technology of Information 9(1), 43–60 (2016)
4. Belousov, K., Dusakova, G.: Analyzer of the rhythmic structure of the text: attribution of texts based on rhythmical patterns. Cifrovaya gumanitaristika: resursy, metody, issledovaniya 1, 49–51 (2017), (in Russian)
5. Bergez, D., Géraud, V., Robrieux, J.J.: Vocabulaire de l'analyse littéraire. Armand Colin (2010)
6. Boychuk, E., Paramonov, I., Kozhemyakin, N., Kasatkina, N.: Automated approach for rhythm analysis of French literary texts. In: Proceedings of 15th Conference of Open Innovations Association FRUCT. pp. 15–23. IEEE (2014)
7. Couranjou, P., Lachambre, B.: Pae stylistique informatique. Computer stylistic, in Le Bulletin de l'EPI 56, 24–35 (1989), (in French)
8. Dubremetz, M., Nivre, J.: Rhetorical figure detection: Chiasmus, epanaphora, epiphora. Frontiers in Digital Humanities 5, 10 (2018)
9. Freyermuth, S.: Poétique de la prose ou prose poétique? le rythme contre le prosaïsme. Questions de style, Vous avez dit prose? pp. 67–80 (2009), (in French)
10. Fromilhague, C., Sancier-Chateau, A.: Introduction à l'analyse stylistique. Bordas (1996)
11. Greene, E., Bodrumlu, T., Knight, K.: Automatic analysis of rhythmic poetry with applications to generation and translation. In: Proceedings of the 2010 conference on empirical methods in natural language processing. pp. 524–533. Association for Computational Linguistics (2010)
12. Kishalova, L.: Analysis of features of rhythmic structure of texts of different styles of the speech. Vestnik Bryanskogo gosudarstvennogo universiteta 1(27), 257–261 (2016), (in Russian)
13. Meschonnic, H.: Critique du rythme. Anthropologie historique du langage. Verdier : coll. "Verdier Poche" (2009), (in French)
14. Niculescu, I.D., Trausan-Matu, S.: Rhythm analysis of texts using natural language processing. In: Proceedings of the 13th International Conference on Human-Computer Interaction RoCHI'2016. pp. 107–112 (2016)
15. Niculescu, I.D., Trausan-Matu, S.: Rhythm analysis in chats using natural language processing. In: Proceedings of the 14th International Conference on Human-Computer Interaction RoCHI'2017. pp. 69–74 (2017)
16. Robrieux, J.J.: Les figures de style et de rhétorique. Dunod (1998)
17. Toldova, S., Azerkovich, I., Ladygina, A., Roitberg, A., Vasilyeva, M.: Error analysis for anaphora resolution in russian: new challenging issues for anaphora resolution task in a morphologically rich language. In: Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes. pp. 74–83 (2016)