

Automated Approach for Rhythm Analysis of French Literary Texts

Elena Boychuk, Ilya Paramonov, Nikita Kozhemyakin, Natalia Kasatkina
P.G. Demidov Yaroslavl State University
Yaroslavl, Russia

elena-boychouk@rambler.ru, ilya.paramonov@fruct.org, enginegl.ec@gmail.com, ninet75@mail.ru

Abstract—Rhythm analysis is widely used for texts in a poetic form to determine the individual style of the author, but rarely used in the analysis of prose due to technical problems and human factor influence. To overcome these issues we propose an automated approach that involves the development and use of specialized software for analyzing French literary prose at various stylistic levels: phonetic, lexical, and grammatical. The methods developed for rhythm analysis and implemented in the computer application cover a variety of the phonostylistic devices: calculation of the length of the rhythmic units, finding assonance, alliteration, rhyme, various repetitions, and others. Efficiency of the approach was proved experimentally by the analysis of rhythmization devices in the novels of four French writers. It was shown that the proposed automated approach allows the researcher to analyze the text 15 times faster than using the manual approach.

I. INTRODUCTION

For defining the individual style of an author, stylistics uses such aspects of analysis as vocabulary, semantics, grammar, text structure, and the preferred stylistic devices. To this variety of methods one can add another one—the analysis of text rhythm. In linguistics, rhythm is usually defined as the regular repetition of similar and commensurable units of speech that perform the structuring, text-forming, and expressive and emotional functions. This aspect is widely used in studying poetic literary works, however, applying this method to the study of literary prose can be problematic [1].

The importance of analyzing the rhythm of a literary text consists of several aspects. First, the analysis of rhythmical characteristics is necessary for determining the authorship of a text. Secondly, such analysis contributes to a deeper understanding of the author's creative method, his artistic conception, and the individuality of his creative work and mastery. Thirdly, the rhythmization degree of a text determines its perception by the reader and the power of its influence on the reader.

Traditionally, rhythm analysis has always relied on quantitative approaches performed manually, which allow for numerous errors in the process.

In this article we offer an approach that involves the development and use of specialized software for the analysis of French literary prose in order to study the functioning of certain rhythmization devices at various stylistic levels: phonetic, lexical, and grammatical.

The use of a specialized application for text rhythm analysis makes it possible to reduce the influence of the

human factor on the research results. The approach also allows researchers to raise their efficiency of working with a text by saving time and facilitating work in such areas as counting words, identifying their sequences, and registering repetitions, which form the text rhythm that are found at each level of the language.

Developing the approach for French texts is primarily caused by the predominance of French in Europe in the Middle Ages (beginning with the 11th century) when French was regarded as the mother-tongue of the upper class. Numerous works of literature were written in French although many texts have now been lost over the years, e.g., the French version of “Tristan et Iseut” by Chrétien de Troyes. As there exist several French versions of the novel (e.g., by Béroul, by Thomas d'Angleterre), determining the authorship when a text is discovered is a crucial task.

The rest of the paper is organized as follows. Section II overviews the existing text processing methods and computer applications that can be used for tasks related to rhythm analysis. The methods that we worked out for detecting the principal rhythmization devices and the application developed on their basis are presented in Sections III and IV respectively. Section V presents the evaluation of the automated approach involving usage of the developed application. It includes a description of our evaluation experiment, the specific features of the functions of the rhythmization devices, which we discovered in the works by 19th century French writers using the proposed approach, and comparison of the time consumed for analysis with and without the developed application. Section VI concludes the paper.

II. RELATED WORK

The principal means of text processing are implemented with the interaction of the quantitative and qualitative analytical methods. From this point of view, text processing consists of text statistical processing and in the singling out of certain groups of textual characteristics and peculiarities at various language levels [2].

Within the sphere of text processing, numerous studies have been dedicated to phonetic, lexical, and grammatical analyses. For instance, L.C. Paquin, J. Beauchemin [3] examine the researcher's role in the computer-assisted process of defining lexical and grammatical text peculiarities. Stressing the interconnection of the morphological, semantic, lexical, and syntagmatic levels, which is mainly defined by the researcher,

the authors come to the conclusion that a fully automated reading of a text is impossible.

The main aspects of computer-assisted text analysis are also described in the study by F.J. Sanchez Perez [4]. The author accentuates the methodology, taking as principal tools data recognition and quantitative processing, as well as the analysis of lexicometric indices (word ranging according to different parameters, i.e., hierarchical, quantitative, alphabetical, and contextual).

There are only a few papers dedicated to the studies of text rhythm with the help of computer applications. Among them, one should point out the study by P. Couranjou and B. Lachambre [5]. Text rhythm analysis is closely connected with stylistics, which involves the use of particular grammatical structures. On the basis of this concept, the French researchers conducted the analysis of sentences for their grammatical structure and rhythm. In pieces of prose by Céline, the number of syllables in the sentences was counted and rhythm curves for different text fragments were drawn accordingly.

In the paper by J.-Ph. Massonnie [6], the rhythmical structure analysis is based on the repetition of the most frequent words. According to it, the text is divided into logical parts.

American researchers E. Greene, T. Bodrumlu and K. Knight [7] analyze poetic text rhythm on the basis of the text metrics in order to translate the text as close to the original as possible (from the point of view of its metrics).

Specialized software can be used for rhythm analysis. Thus, *Alceste*¹ developed by IMAGE in co-operation with Centre National de la Recherche Scientifique (CNRS) conducts a detailed analysis of the text vocabulary grouping the words according to their lexical and semantic characteristics and to their frequency of use. According to the result, the text is divided into parts. The method allows to find the most valuable word combinations from the point of view of their meaning. This software can be used for finding the key words in a text and dividing it into logical parts.

In the area of rhythm analysis in Russian, English and French, *Rhymes*² by N. Ketsaris assists in processing the text in order to find rhymes. On the basis of the dictionary of Russian grammar by A. Zalisnyak, the rhymes are sorted alphabetically into their individual qualities or parts of speech. Limits for the rhyme list can be set according to the part of speech or the number of syllables. Besides, there is an option for finding the synonyms and antonyms to the set word. The vocabulary corpus is based on the dictionary of synonyms by A. Yevgenyeva.

The metrics examining application *Ritminme*³ by S. Zigulya analyzes the metric structures of poetic texts. Recognizing the stressed and unstressed syllables, it determines the metric scheme. Besides, rhymes for the given word are found.

French-based *Tropes* developed by P. Molette & A. Landré [8] is grounded on the works by R. Ghiglione, a psychologist and discourse specialist. The method is aimed at delineating

the lexico-semantic field of the text, analyzing its morphological structure, and determining its functional type (narrative, descriptive, speculative, argumentative). With the assistance of this method, lexico-semantic groups are recognized and their frequency of use in the text is estimated. Lists of verbs, nouns, adjectives, adverbs, conjunctions, and pronouns with various meanings within the fragment are given. The verbs are classified according to their semantic type.

Each of the tools presented in this section is oriented towards a certain set of rhythm analysis means. This article is aimed at analyzing the rhythm at the phonetic, lexical, and grammatical levels of implementation; within this concept, the last method, *Tropes*, is the most valuable. It allows, first, to work with French texts; second, to determine the number of repetitions in the text (of words belonging to different parts of speech).

However, the stated method is not sufficient for determining the degree of the text rhythmicality, as rhythm is a complex phenomenon implemented through a range of linguistic devices at different levels. The approach offered in this article makes it possible to assess the rhythm both from multiple points of view of grammar, phonostylistic (assonance, alliteration, rhyme, rhythmic units), lexical, and grammatical (repetitions of various types—epanalepsis, reduplication, and anadiplosis) rhythmization devices.

III. METHODS

Within the phonetic aspect, we offer the analysis of such means of rhythm implementation as the length of the rhythmic units (equality or sequence of syllables), assonance, alliteration, and rhyme.

Within the lexical aspect, analysis of the repetition of words is conducted depending on the frequency of their occurrence. Repetitions of synsemantic words are not subject to analysis.

The grammatical aspect of rhythmization analysis is based on studying the morphemic structure of words (derivation), coordinated words marked by certain graphic symbols, sentences of the same communicative type (interrogative, exclamatory, aposiopesis), anaphora, epiphora, symploce, anadiplosis, reduplication, epanalepsis, climax, and polysyndeton.

A. The length of the rhythmic units

For calculation of the length of the rhythmic units, we consider syllabification rules, reading rules, sound combinations, diphthongs which should be perceived as one syllable (e.g., *iè*, *ieu*, *eai*, etc.), as well as the various positions of consonants whose pronunciation depends on their context. Besides, while analyzing the length of the rhythmic units, the traditional short forms (such as *Mme*, *M.*, etc.) were taken into consideration.

The pronunciation rules for the dropped “e” are difficult as the rules are not fixed and the pronunciation depends on the functional style, the tempo, and many other factors including the individual manner of reading. In this research, the pronunciation rules for the dropped “e” in the position of two consonants before the third consonant were taken as a basis.

Taking into account the syllabification rules and the numerous exceptions, we developed an algorithm for singling

¹<http://www.image-zafar.com/en/alceste-software>

²<http://rifmovnik.ru>

³<http://www.ritminme.ru>

out the rhythmic units of the text. As it is difficult to make the singling out of rhythmic groups comply with the general phonetic rules (this requires a complete analysis of the syntax of the sentence), we used the division rules with the help of the punctuation marks (commas, colons, semicolons, dashes) as well as coordinate and subordinate conjunctions.

There is no doubt that such division is not quite precise. However, at this stage, we can regard it as division into certain rhythmic units facilitating the study of the text rhythm, not division into rhythmic groups. By dividing the text into rhythmic units and counting the syllables in them, it is possible to see their equality, progression, or sequence, which is closely connected with the text rhythm: the units with the equal number of syllables and the units forming a sequence with a difference of one syllable are found to be the most rhythmical.

B. Assonance

Historically, Romance assonance served as a transition from the complete absence of rhyme to the beginnings of assonance, and then to rhyme. There is no consensus among French linguists in defining assonance. Some sources point to the understanding of assonance as the repetition of the vowel manifested in the final stressed syllable; others include in the concept of assonance the repetition of vowel sounds, regardless of their position and stress.

The first position is justified by the fact that vowel sound repetitions are more noticeable when they are formed by stressed vowels. However, in this case it is quite difficult to distinguish the phenomena of internal rhyme and assonance: "...j'ai moi, Colas Breugnon, bon garçon, Bourguignon, rond de façons et du bedon" (R. Rolland, "Colas Breugnon").

However, the process of analyzing assonance as a means of rhythmization [9] revealed instances of vowel repetition in the unstressed position within certain rhythmic units: "Un peu de pluie tombait; au loin, les flocons de brume montaient, comme de grands fantômes" (G. Flaubert, "Bouvard et Pécuchet"). The example given combines the repetition of two types of sounds: in the stressed and unstressed positions.

Actually, assonance in French is defined as the repetition of rhyming vowels both at the end of the word in the stressed position (and this is its predominant position, as it reveals rhyme the most fully), and in the unstressed position. The main difference between assonance in the final (stressed) position and rhyme lies in the fact that assonance is formed by the coincidence of vowel sounds only, without the participation of the preceding or following consonants.

In our view, instances of repetition of unstressed vowels in the middle of the words should not be excluded, as they influence the rhythmic structure of the text and the way the text is perceived by the reader. So, in this research, the repetition of sounds in both the stressed and unstressed positions within the rhythmic unit is regarded as assonance.

In order to reveal the repetition of vowel sounds in text fragments, we formulated the rules which allow the program to differentiate between mute and pronounced vowels depending on their position and combination. The total number of instances of use is given for every symbol to enhance the

demonstrability of alliteration and assonance in working with large text fragments.

Repeated vowels are detected on the basis of understanding assonance as the repetition of both stressed and unstressed vowel sounds, which is connected with the researchers' growing interest in the rhythm of prose where assonance serves as a means of rhythmization and belongs to the sphere of general literary criticism, outgrowing prosody. In this respect, the program allows to single out as units of assonance not only separate sounds but also groups of sounds which are close to each other phonetically. For instance, the symbol "a" covers the sounds [a], [a], [wa], [ua], [ya], [aj], [ja]; "e" stands for the sounds [ej], [je], [e], [εj], [jε], [yε], [uε]; etc.

C. Alliteration

As in the case of assonance, in order to reveal the repetition of consonant sounds in text fragments, we formulated the rules which allow for differentiating between mute and pronounced consonants depending on their position and combination.

D. Rhyme

Rhyme is rather widely used in prose. The abundance of rhymed words in the text is a manifestation of the writer's style and is also conditioned by the author's desire to attract the reader's attention to a particular character, his personality and peculiarities of speech.

The quality of the rhyme depends on the number of coinciding consonant sounds in the word endings. Within the suggested method, rhymes are classified into the following types:

- 1) Poor rhyme (only the final vowel in common: canaux—vaisseaux);
- 2) Insufficient rhyme (two phonemes in common, one of them being the stressed vowel: mystérieux—yeux);
- 3) Rich rhyme (three and more phonemes in common, one of them being the stressed vowel: ensemble—ressemble).

The search for the three types (poor, insufficient, rich) is based on finding identical endings and is accompanied by their counting. In the case of a poor rhyme, only the final stressed vowels coincide; the coincidence of two phonemes including the final stressed vowel presents an insufficient rhyme; a rich rhyme is characterized by the coincidence of three and more phonemes (including the final stressed phoneme) [10]. The researcher's role consists in finding closely situated repetitions of similar endings which form the rhyme.

On hearing the second rhyme, we unconsciously revive the image of the first rhyming word, thus the inner connections of the content are reinforced and clarified externally. If the distance is too big, the human mind might not perceive the rhyme.

E. Grammatical aspect

Studying the devices listed at the beginning of this section as a means of rhythmization in prose is not a mainstream phenomenon in linguistics. At this language level, the functioning of text rhythm was examined in the works by M. Croll

[11], N.R. Tempest [12], G. Dessons [13], and some others. Repetitions of lexical units and syntactic constructions, which are structurally connected with them, with semantic variations, were studied to a greater or lesser degree in the works by those authors.

The above-mentioned grammatical means are based on repetition, which has the following structure: “the original element + repetitions of the original elements”. Depending on the device, the elements can occupy various positions towards each other in the rhythmic unit: initial, final, medium, contact, and non-contact. Thus, anaphora and reduplication are characterized by the initial position in the rhythmic unit; epiphora—by the final position; symploce—by the initial and final ones; and chiasmus—by the crossed position of repeated elements [14]. For other devices, their position towards each other is more important, e.g. for anadiplosis, gradation, epanalepsis, and coordinated words.

To achieve a greater degree of rhythmization with the help of the stated devices the following conditions are required: (1) presence of a number of repeated elements; (2) the smallest possible distance between the original element and the repetitions; (3) high frequency of occurrence in the text.

The grammatical rhythmization devices are closely connected with the phonetic level, namely, with the suprasegmental expressive means. This is conditioned by the fact that syntactic constructions based on syntactic parallelism share the same intonation patterns, as well as coordinated words, repeated inverted constructions, and sentences belonging to different communicative types do. The influence of the phonetic aspect enhances the role of the grammatical devices in the perception of text rhythm. Nevertheless, it is necessary to estimate the frequency of occurrence both for the grammatical means of rhythmization and for the phonetic (segmental, phonostylistic) and lexical ones for more complete mental and emotional text perception. The program determines the frequency of occurrence for the above-mentioned devices, which makes it possible to estimate their periodicity and, consequently, the rhythmic structure of the text. The use of certain rhythmic devices in a prosaic literary work, in its turn, allows us to evaluate the peculiarities of the author’s language, and his idiosyncratic style.

IV. APPLICATION FOR RHYTHM ANALYSIS

A. Overview

We developed the application for rhythm analysis using Qt Framework and Qt Quick technology. It can be executed on a variety of desktop platforms including MS Windows, GNU/Linux, and Mac OS X.

The application allows to analyze the rhythm of French literary texts by applying a number of methods grouped according to the aspect of analysis: phonetic, lexical, and grammatical.

The phonetic aspect involves the search for alliteration, assonance, three types of rhyme and the measurement of rhythmical units. The rhythmical devices are identified in accordance with French pronunciation and syllabification rules. In addition, the authors of the article worked out lists of letter combinations for determining the length of rhythmical units,

assonance, alliteration, rhyme, and lists of exceptions. The summary on these lists is presented in Table I.

TABLE I. LISTS OF TEXT ELEMENTS USED BY THE APPLICATION

List description	Number of elements	Methods	Examples
List of French vowel symbols	20	Determining the length of rhythmical units, finding assonance, finding rhyme	i, o, ù, à, è, é
List of vowel diphthongs	130	Determining the length of rhythmical units	ui, ai, oui, ue, ie
List of consonant diphthongs	18	Determining the length of rhythmical units	gg, ch, gn, ll
List of conjunctions	93	Determining the length of rhythmical units, finding polysyndeton	mais, sinon, comme, lorsque
List of sysemantic words	280	Determining the frequency of use	un, une, des, la, le
List of monosyllabic words	19	Determining the length of rhythmical units, finding assonance	que, ques, je, le
List of words in which the ending ‘ent’ is pronounced	174	Determining the length of rhythmical units	arpent, auvent
List of words in which the ending ‘ent’ is not pronounced	128	Determining the length of rhythmical units	affament, afferment, affirmant
List of poor rhyme endings with exceptions	12	Finding rhyme	[a]: a,ah,as
List of insufficient rhyme endings with exceptions	147	Finding rhyme	[ab]: abe, abes, abent
List of rich rhyme endings with exceptions	80	Finding rhyme	[able]: able, ables

The lexical aspect of the analysis is carried out by detecting of the frequently used words.

The grammatical aspect involves the search for coordinated words, sentences of the same communicative type, as well as repetitions of various kinds.

Fig. 1 depicts the main menu of the application. The user chooses the module and then the method required for the analysis. After that the text to be analyzed is entered into the corresponding field. In the case of searching for alliteration, assonance and rhyme, the user should additionally select the sound for the detailed analysis and for receiving the results according to the sample.

The phonetic aspect presents the greatest interest for analysis, as it is primarily at the phonetic level that rhythm is formed. Within the phonetic aspect, the search is conducted symbol by symbol, unlike the grammatical aspect wherein the texts are analyzed sentence by sentence or the sentences are recognized as spaces, words, and punctuation marks. Details of implementation of the analysis methods are given in the following subsections.

B. The length of rhythmical units

The function of measuring the length of rhythmical units allows to demonstrate the rhythmization in a text visually.

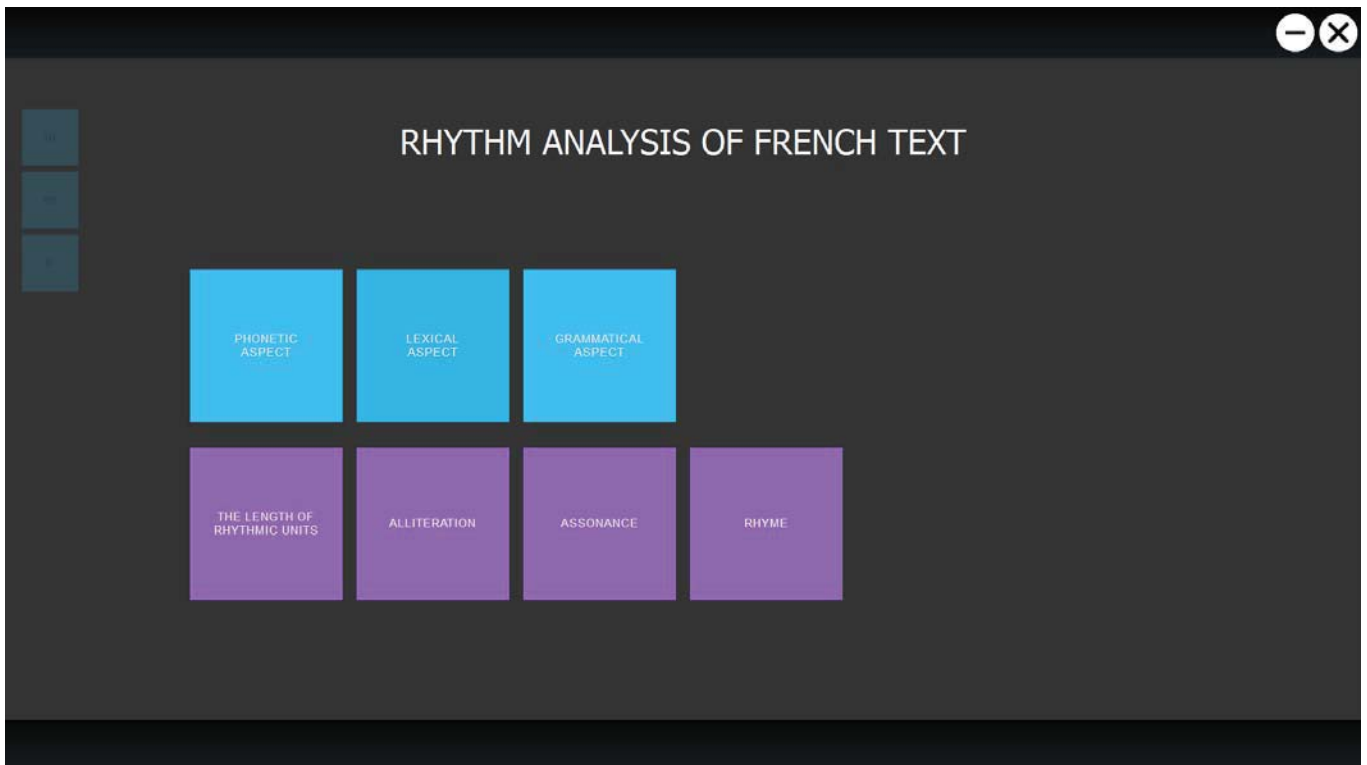


Fig. 1. Main menu. The top line shows aspects of analysis. According to the aspect chosen, the bottom line changes, revealing the analysis methods for the aspect

The first step consists in dividing the text under study into rhythmical units. The borders are identified by punctuation marks and by co-ordinating and subordinating conjunctions. The second step consists in counting the number of syllables in each rhythmical unit. Within this step, the letters signifying French vowels are analyzed according to the following rules:

- 1) Whether there is a diphthong, according to the list of diphthongs;
- 2) Whether the word is monosyllable;
- 3) If the letter “e” is found it is checked according to the special reading rules for this letter;
- 4) The other sounds are analyzed in accordance with the corresponding rules and exceptions.

The result consists in determining whether the vowel is a silent one.

The final step involves counting the number of pronounced vowels in the rhythmical unit. The group is colored according to the resulting number of syllables which is marked at the end of it. This makes it easier for the researcher to see the analysis results.

The example of the application output is presented in Fig. 2. Groups with the same number of syllables are marked with the same color. Each group is followed by the number of syllables in square brackets. Such representation allows to examine the text visually and easily identify the units with the identical number of syllables and those in which the number of syllables differs by one or two syllables.

C. Alliteration and assonance

The proposed method of searching for alliteration and assonance includes counting the sounds in every word (vowels are counted for finding assonance, consonants for alliteration), with the subsequent identification of the sounds requested by the user in the text.

The first stage consists in processing the whole text and displaying the list of sounds with the number of their instances of use in the text. This allows the researcher to single out the most frequently used sounds.

At the second stage of the analysis, the search for a particular sound chosen by the user is conducted. The instances of use are highlighted, which allows the user to recognize alliteration and assonance. An example of a search for alliteration is given in Fig. 3.

D. Rhyme

For every type of rhyme there is a list of sounds with the corresponding endings as well as additional rules and exceptions (for example, the words with the similar endings that are not suitable because of pronunciation rules). According to the list of endings for the chosen rhyme type, analysis of the endings of the words in the text is conducted.

As a result, the list of sounds is presented with the number of instances of their use indicated. At the second stage, the endings finishing in the chosen sound are highlighted.

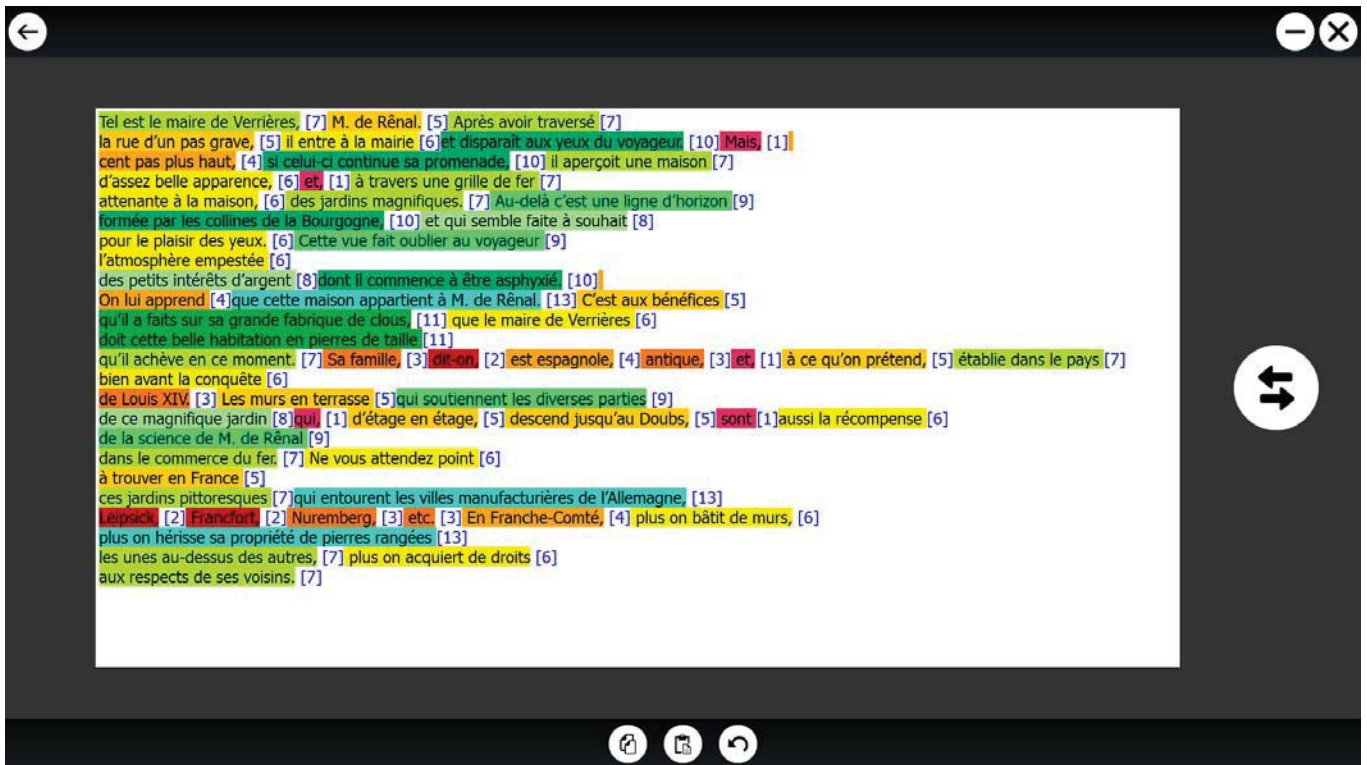


Fig. 2. Results of measuring the length of rhythmical units. The text is divided into rhythmical groups, at the end of each group the number of the pronounced syllables is displayed. Each rhythmical group is marked by color according to the number of pronounced syllables

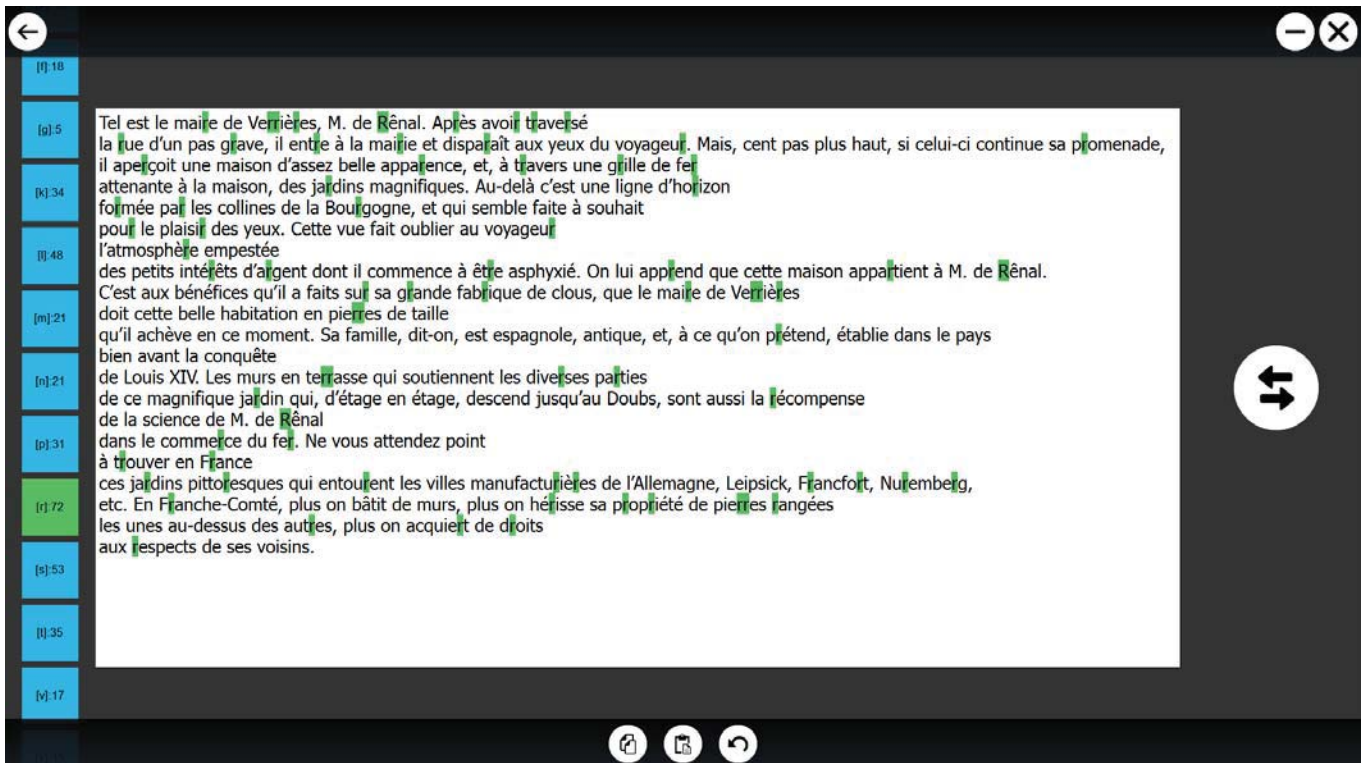


Fig. 3. Results of a search for alliteration in a text. On the left, there is a column containing all the consonant sounds found and the number of instances of their use in the text. At the second stage, the user chose the sound [r], so all the instances of its use are highlighted

E. Frequency of word use

Analysis of the most frequently used words is essential for studying the rhythm of a literary text. Lexical, semantic, and thematic repetitions form semantic fields which help to construct the topical structure of the text. Word repetition forms “a thematic network of the text and is connected with its content”; besides, “if it is consistent, it emphasizes the point of view of the character or the narrator” [15]. Within this method, the number of word repetitions in the text is defined. Synsemantic words are identified by the list and excluded from the results of the search.

F. Sentences of the same communicative type

In order to find the sentences belonging to the same communicative type, the text is divided into sentences and their type is defined (exclamatory, interrogative, and the sentences with ellipsis). Then the sentences are examined for the coincidence of their communicative type (narrative sentences are outside the scope of our research within this aspect).

G. Word repetitions

Within the aspect of grammar, word repetitions of different kinds are examined as a means of rhythmization.

For identifying anaphora, the initial words in the parts of a complex sentence or neighboring sentences in a text are compared, e.g., “*Âge* où tout est luisant, où tout scintille et flambe! *Âge* de force joyeuse dont personne ne profite” (from O. de Balzac, “*Père Goriot*”).

For identifying epiphora, the final words in the parts of a complex sentence or neighboring sentences in a text are compared, e.g., “*Nous* voudrions vous voir bastant comme notre père, et point dans la *fosse!* La *fosse!* Nous y sommes toujours assez tôt, dans la *fosse.*” (from O. de Balzac, “*La peau de chagrin*”).

The search for symploce is conducted in two stages. The first step involves the dividing of the sentences into parts according to the punctuation marks. The second step is to compare the initial and final words in every sentence part and in the whole sentence, e.g., “*J’ai* gagné une *bataille*, se dit-il aussitôt qu’il se vit dans les bois et loin du regard des hommes, *j’ai* donc gagné une *bataille!*” (from Stendhal, “*Le rouge et le noir*”).

For finding reduplication, the beginning of each sentence is examined for the repetition of a word after a comma, exclamation, or question mark, e.g., “*Arrête! arrête!* la période n’est pas accomplie.” (from G. Flaubert, “*Bouvard et Pécuchet*”).

The search for anadiplosis consists of the comparison of words at the junctures of the sentences and their parts, e.g., “*Maréchal* avait été *blond, blond* comme Jean.” (from G. de Maupassant, “*Pierre et Jean*”).

The search for epanalepsis involves the search for identical words within a sentence part or a sentence. The identical words standing next to each other are excluded, e.g., “*Oh! ton amour, Raphaël, ton amour* vaut le monde.” (from O. de Balzac, “*La peau de chagrin*”).

Identifying the climax should be mentioned specially. In order to find it, repeated words with gradation-forming adverbs

are searched for, e.g., “*Une connaissance approfondie et trop approfondie* des *Saintes Écritures.*” (from Stendhal, “*Le rouge et le noir*”).

H. Polysyndeton

For identifying the use of this device, the list of coordinating and subordinating conjunctions is used. With the help of it, conjunctions are recognized and highlighted in every sentence.

I. Coordinated words

In order to find coordinated words, the text is divided into sentences. In each sentence constructions that match the pattern “word—comma—word—comma—word—...” with three and more coordinated words are identified and highlighted.

This method mainly focuses on constructions of the following type: “*Enjouée* jadis, expansive et tout aimante, elle était, en vieillissant, devenue (à la façon du vin éventé qui se tourne en vinaigre) d’humeur *difficile, piaillarde, nerveuse.*” (from G. Flaubert, “*Madame Bovary*”). Such constructions are the most expressive rhythmically, as they are accompanied by a special intonation of enumeration and quite often by equalities or sequences of syllables in rhythmical units.

There is no doubt that in searching for coordinated words using the proposed method, a number of inaccuracies are possible, namely those connected with identifying such units separated by commas as forms of address, proper names, adverbs, participles and conjunctions introducing subordinate clauses, e.g.: “*Vers* quatre heures du *matin, Charles, bien* enveloppé dans son manteau, se mit en route pour les Bertaux.” (from G. Flaubert, “*Madame Bovary*”). Another example: “...car c’est à *vous, surtout, que* j’en dois la publication.” (from G. Flaubert, “*Madame Bovary*”).

However, such instances are not numerous and can be manually filtered out by the researcher.

V. EVALUATION

The application was tested in co-operation with 4th-year students of the Yaroslavl Teacher Training University named after K. Ushinsky (Faculty of Foreign Languages). The students were asked to use the program to determine and analyze the use of the above-mentioned rhythmic devices in the works of four French writers (Stendhal, H. de Balzac, G. Flaubert, G. de Maupassant) in order to define the individual peculiarities of their styles and to estimate the rhythmicality of the texts in connection with the frequency of use of the rhythmic devices. The following works were studied with the help of the application:

- Stendhal (1783–1842): “*Le rouge et le noir*” (1830), “*La Chartreuse de Parme*” (1839), “*Lucien Leuwen*” (1835);
- H. de Balzac (1799–1850): “*Les Chouans*” (1829), “*La Peau de chagrin*” (1831), “*Le Colonel Chabert*” (1832), “*Eugénie Grandet*” (1833), “*Le Père Goriot*” (1835), “*Illusions perdues*” (I, 1837; II, 1839; III, 1843);

TABLE II. TYPICAL LENGTH RANGE OF THE MOST FREQUENT RHYTHMIC EQUALITIES AND SEQUENCES

Author	Title	Equalities	Sequences
Stendhal	Le rouge et le noir	4–5	3–7
	La Chartreuse de Parme	4–6	4–7
	Lucien Leuwen	4–8	1–7
H. de Balzac	Le Père Goriot	2–5	2–7
	Le Colonel Chabert	1–7	2–8
	Les Chouans	3–8	1–8
	La Peau de chagrin	2–6	3–8
	Illusions perdues	3–8	3–7
	Eugénie Grandet	4–7	2–6
G. Flaubert	Madame Bovary	2–5	2–7
	Bouvard et Pécuchet	4–7	2–6
	Education sentimentale	3–6	2–7
	Salammbô	3–8	2–7
G. de Maupassant	Une vie	2–6	2–6
	Bel ami	2–6	2–6
	Mont-Oriol	2–6	2–6
	Pierre et Jean	2–4	2–6
	Fort comme la mort	4–8	3–7
	Notre coeur	3–6	2–6

- G. Flaubert (1821–1880): “Madame Bovary” (1857), “Education sentimentale” (1869), “Bouvard et Pécuchet” (1881), “Salammbô”;
- G. de Maupassant (1850–1893): “Une vie” (1883), “Bel ami” (1885), “Mont-Oriol” (1887), “Pierre et Jean” (1888), “Fort comme la mort”, (1889), “Notre coeur” (1900).

The experiment revealed the following rhythmic peculiarities of the texts by the 19th century writers:

1. The analysis of Stendhal’s works revealed a low occurrence frequency of the rhythmic devices defined within the phonetic aspect, as well as their low density within the rhythmic units. This makes it possible to characterize the texts as the least rhythmical in comparison with the other authors’ works. The average length of rhythmic units is 4–7 syllables (see Table II). The range of using rhyming words within a rhythmic unit is 2–4 rhythmic groups. The interposed groups without rhyming sounds are called gaps. Stendhal’s novels are characterized by larger gaps than those of the other authors: 0–4 groups, which points to the low density of the use of rhyme. Within the phonetic aspect, assonance is most widespread (see Table III). The range of its use depends on the number of rhythmic groups in which it is revealed. Stendhal’s works contain the smallest number of rhythmic groups (up to 10), united by assonance. Alliteration is mainly formed by the repetition of the following 10 sounds: [r], [s], [f], [m], [b], [k], [d], [ʒ], [ʃ], [l]. At the level of grammar, the rhythmization of the text is characterized by the use of constructions with coordinated words (typically including no more than 2–4 elements), lexical anaphora and climax (see Table IV).

TABLE III. FREQUENCY OF USE OF PHONOSTYLISTIC RHYTHMIZATION DEVICES

	Stendhal	H. de Balzac	G. Flaubert	G. de Maupassant	Total
Rhyme	427	531	617	601	2176
Assonance	452	397	378	443	1670
Alliteration	472	493	430	534	1929

2. H. de Balzac’s novels are considered to be in many aspects more rhythmical as compared to those of Stendhal.

TABLE IV. FREQUENCY OF USE OF THE MOST WIDESPREAD GRAMMATICAL RHYTHMIZATION DEVICES

	Stendhal	H. de Balzac	G. Flaubert	G. de Maupassant	Total
Derivation	89	101	71	334	2782
Lexical anaphora	695	862	502	723	2782
Coordinated words	437	744	1094	2100	4375
Anadiplosis	41	143	56	265	511
Epanalepsis	110	157	166	489	922
Climax (gradation)	154	51	80	285	570

The author’s works are characterized by two- and three-group equalities, as well as by sequences with an equality included. Balzac’s novels have the largest range for using rhyming words within this or that rhythmic unit, which is two to six rhythmic groups. The gap is rather small and makes zero to two rhythmic groups. Assonance-forming vowels are repeated within 10–15 groups. Compared to Stendahl’s novel, this is a higher rate. Alliteration is found to be the most popular device in Balzac’s works and as in the case of Stendahl, is mainly formed by the repetition of the following 10 sounds: [r], [s], [f], [m], [b], [k], [d], [ʒ], [ʃ], [l]. At the same time, Balzac’s novels demonstrate a higher density of alliterating sounds used within one rhythmic unit. The grammatical aspect is characterized by the frequent use of structures with anaphora, and also by the use of coordinated attributes expressed by adjectives or participles and coordinated predicates expressed by verbs (2–4 components on average).

3. At the segmental sublevel, the novels by Flaubert are characterized by 4-syllable equalities including two to five groups, and also two-group rising equalities. The range for using rhyming words within this or that rhythmic unit is two to three rhythmic groups which makes the smallest number compared to the other authors’ works. However, the gaps are the smallest—zero to two rhythmic groups. The most widespread phonetic device is the rhyme. Flaubert’s works demonstrate the largest number of rhythmic groups united by the repetition of assonance-forming vowels (up to 20). The texts show the smallest variety of alliterating sounds: [r], [s], [b], [d], [m], [ʒ], [k], [l], [ʃ]—9. Repetitions of several consonants within one rhythmical unit are not typical. From the point of view of the grammatical aspect, Flaubert’s novels are characterized by the use of coordinated attributes and objects expressed by adjectives, participles and nouns (from 2 to 5).

4. As for Maupassant’s works, they are characterized at the segmental sublevel by equalities which include two to five syllables and are used both independently and within sequences. Owing to the smallest quantitative difference between the equalities and sequences, and the most frequent use of equalities as compared with the other authors’ novels, the works by Maupassant are the most rhythmical. The most frequently used devices are rhyme and assonance, though alliteration stands very close. Alongside with Balzac’s novels, works by Maupassant demonstrate the greatest range for using rhyming words—two to six rhythmic groups with the minimum gap of 0–2 rhythmic groups. The range of assonance in the texts by Maupassant makes fifteen rhythmic groups, the difference with Flaubert’s texts being five groups. The high frequency of using different types of rhyme is also evident.

This brings us to the conclusion that in comparison with the works by the other authors, those by Maupassant reveal the greatest degree of rhythmization. Alliteration is presented by repetitions of 12 consonants: [s], [r], [ʒ], [l], [m], [b], [f], [ʃ], [k], [d], [z], [g]. High frequency, high density, and repetitions of a number of consonants within one rhythmic group are typical of this author's works. As for the grammatical aspect, the prevailing devices are derivation (at the morphological sublevel), constructions with coordinated adverbial modifiers, predicates expressed by verbs in a simple form, attributes expressed by adjectives and nouns, and also constructions including a number of coordinated words groups (from 2 to 17). One should also point out the use of derivation, epanalepsis, anadiplosis, and climax.

It should be noted that the suggested approach considerably reduced the time required for working with the texts. The students who conducted the analysis according to the stated parameters manually spent 15 times as much time as those who used the automated approach.

VI. CONCLUSION

In this paper we presented an automated approach for the complex analysis of rhythm in French literary texts. Our approach is aimed at facilitating rhythm analysis according to different parameters and at raising the productivity of research by reducing the time it requires. The efficiency of the approach was proved experimentally by the analysis of rhythmization devices in the novels of four French writers.

A high or low frequency of appearance of certain rhythmization devices or groups makes the authors' works recognizable for the reader. For example, using the suggested approach in our research, we determined that G. de Maupassant's novels are the most rhythmical in comparison with the novels of the other authors analyzed. This is proved by a high frequency of use of rhythmical means (assonance, alliteration, rhyme, reiteration), while in the works by G. Flaubert rhythmical repetitions are quite infrequent. In determining the authorship of a text, getting such information is important when supported by methods of detailed structural analysis of the rhythm (including the analysis of the visual structure and plot structure of the text).

The value of the proposed approach is also determined by the possibility of using its results in further studies devoted to the issues of rhythmic structure of prosaic texts and a number of other issues connected with the complex analyses of literary texts. For example, rhythm analysis with the help of the suggested approach facilitates and accelerates working out the typology of rhythmic structures created by rhythmic devices at various language levels; optimizes the comparative analysis of the rhythm in the texts belonging to different epochs, authors, styles, and literary trends; and the study of the influence of the

functional types on the text rhythm. It also facilitates finding interdependence between the stated rhythmic devices and the acoustic parameters in audio books, contributes to the study of the phenomena of textual arrhythmia and rhythmic polyphony, and the golden section rule in rhythmic schemes of literary texts.

The approach can be used both in the professional activity of philologists and in practical classes teaching the analysis and interpretation of French texts.

ACKNOWLEDGMENTS

The article was published with financial support by the project No. 549 of P.G. Demidov Yaroslavl State University within State Assignment for Research.

REFERENCES

- [1] S. Freyermuth, "Poétique de la prose ou prose poétique? Le rythme contre le prosaïsme" [Prose poetics or poetic prose? Rhythm against prose], *Questions de style*, vol. 6, 2009 (in French).
- [2] V. Komis, Ch. Depover, Th. Karsenti, *L'usage des outils informatiques en analyse des données qualitative* [The use of computer tools in analysis of qualitative data]. Canada, 2013 (in French).
- [3] L.-C. Paquin, J. Beauchemin, *Apport de l'ordinateur à l'analyse des données textuelles* [Contribution of computer analysis of textual data]. Montréal, 1989 (in French).
- [4] F. J. Sanchez Perez, "Qu'est-ce que l'analyse relationnelle informatique des textes?" [What is the relational analysis computer texts?], in *a Revue Informatique et Statistique dans les Sciences humaines XXIX, 1 à 4*. Paris, 1993 (in French).
- [5] P. Couranjou, B. Lachambre, "P.A.E. Stylistique informatique" [Computer stylistic], in *Le Bulletin de l'EPI*, vol. 56. Paris, 1989.
- [6] J.-Ph. Massonnie, *Analyse informatisée des textes* [Computerized text analysis]. Franche-Comté, 1990 (in French).
- [7] E. Greene, T. Bodrumlu and K. Knight, *Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation*. EMNLP, 2010.
- [8] P. Molette, "De l'APD à Tropes: comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste" [Propositional Discourse Analysis to Tropes: evolution of content analysis in general semantic classification], in *Psychologie Sociale et Communication*. Tarbes, 2009 (in French).
- [9] E.I. Boychuk, "Realizacia segmentnih i fonostilisticheskikh sredstv ritmizacii v romane G. de Mopassana "Mili drug"" [The realization of segment and phonostylistic devices of rhythm in the novel by G. de Maupassant "Bel ami"], in *Vestnik of Tambov University*, vol. 1. Tambov, 2014 (in Russian).
- [10] B. Wenk, "Is French really syllable-timed?", in *Journal of Phonetics*, vol. 10/2. USA, 1982.
- [11] M. Croll, Style, "Rhetoric and Rhythm", in *Essays by Morris W. Croll*. New Jersey, Princeton: Princeton University Press, 1966.
- [12] N.R. Tempest, *The Rhythm of English Prose*. Cambridge: University Press, 1930.
- [13] G. Dessons, "Traité du rythme. Des vers et des proses" [Rhythm treatise. Verse and prose]. Paris: Dunod, 1998.
- [14] L. Jeffries, D. McIntyre, *Stylistics*. Cambridge: University Press, 2010.
- [15] E. Brown, *Rhythm in the Novel*. Toronto: University of Toronto Press, 1967.