Rhythm analysis of Romanian texts

Veis Onur-Șener

University Politehnica of Bucharest
313 Splaiul Independentei,
Bucharest, Romania
onur.veis@yahoo.com

Ștefan Trăușan-Matu

University Politehnica of Bucharest
313 Splaiul Independentei,
Bucharest, Romania
and
Research Institute for
Artificial Intelligence
and
Academy of Romanian Scientists
stefan.trausan@upb.ro

ABSTRACT

The paper presents an implemented interactive application for the analysis of the rhythmicity of texts in the Romanian language, using methods based on artificial intelligence. For the text analysis, are used three mathematical methods developed by Solomon Marcus, Mihai Dinu, and Vasile Vasile. The application was used for analyzing four genres of text: poems, prose, speech, and scientific text.

Author keywords

Rhythm; accent; syllable; natural language processing; artificial intelligence

ACM Classification Keywords

I 2.7 Natural Language Processing. Text analysis

DOI: 10.37789/ icusi.2024.16

INTRODUCTION

Rhythm is generated in text by word repetition [1] or by the periodic and symmetrical succession of stressed and unstressed syllables in a text [2]. Rhythm plays an important role in enhancing language and conveying messages, having also a rhetorical role [3]. Rhythm analysis with natural language processing tools can be used for text classification [4] and, consequently, for author identification. Moreover, the musicality provided by repetition makes the text more pleasant to read and easier to understand.

In Romanian, similarly with some other languages, rhythm is determined by considering the division of words into syllables and the stressed (accentuated) syllables. Although syllable division is a simple task for humans, being the sound or group of sounds pronounced in a single jaw movement, implementing an algorithm for this task is much more challenging. There are rules for dividing words into syllables, but these cover a few cases, leaving many exceptions.

The accent is the highlighting by phonetic means of a syllable in a word [5]. In Romanian, most words have the accent on the last two syllables. There are no criteria that can be implemented in a classical accent determination algorithm.

To solve the above two problems, the RoLEX dictionary [6] was used, which includes information about approximately 330,000 words. It contains information related to word form, part of speech, syllable division, lexical stress and phonetic transcription. RoLEX is based on two dictionaries of a small size RoSyllabiDict and MaRePhor and an algorithm that uses neural networks to make predictions.

Mathematical formalisms were proposed for analyzing rhythm in Romanian language by Solomon Marcus [7] and Mihai Dinu [8]. Vasile Vasile implicitly analyzed rhythm through the euphonic repetition of groups of vowels in the poems of Mihai Eminescu [9]. Only the first two methods of rhythm analysis have been successfully implemented only for English [1] because at that time there were not available all the needed reliable natural language processing tools.

The aim of the research presented herein is to implement all the above three methodologies to analyze the rhythm in the case of Romanian texts. It was desired to implement an algorithm as precise as possible for analyzing the rhythm of large texts. This will allow a faster and more efficient understanding of the structure of the texts, thus automating a time-consuming process if done manually. Using the methods described was analyzed the rhythmic features in four genres of texts (poems, prose, speech, and scientific text).

The paper continues with a state of the art about rhythm, about Romanian digital dictionaries and three methods for rhythm analysis. The next sections describe the implementation of the three methods, the results, and the conclusions.

STATE OF THE ART

The history of rhythm

The concept of rhythm has been studied since ancient times. Aristoxenos from Tarent believed that "Rhythm is an arrangement determined by times". During the Middle Ages and the Romantic period, rhythm continued to be redefined. In the 20th century, interest in rhythm increased significantly, thus generating many studies on it.

Miron Costin is considered the initiator of versification studies in Romanian literature. In the preface to his philosophical poem "Poemul Lumii", he defines elements such as measure, rhyme and hiatus. According to Nicolae Cartojan, this is "the first of its kind in our literature" [14, p. 217].

Another pioneer in the study of versification was the monk Macarie, who in his grammar dedicated a part to "the craft of the verse maker of poetics". It marks the equality between Latin and Romanian versification, classifying vowels into long, short and common [14, p. 217].

The first international congress dedicated to rhythm took place in Geneva in 1926. The Romanian mathematician Matila C. Ghyka and the researcher Pius Servien-Coculescu were among the first to contribute to the understanding of rhythm from a mathematical perspective [14, p. 209].

Pius Servien, in the work "Les Rythmes comme introduction physique à l'Esthétique", defines rhythm as "perceived periodicity". He believes that rhythm acts by deforming the usual flow of time, highlighting the periodicity of natural phenomena [14, p. 211].

The history of digital dictionaries for the Romanian language

An algorithm was developed in ASSIRIS assembly language on a FELIX C-256 computer, using a database of 2,000 words for generating paradigms of Romanian words. The result was the Morphological Dictionary of the Romanian Language (1973–1974), with 2,058 words and an automaton for inflection and lemmatization [10, p. 3].

The TEZAROM Lexicon (1992–2010) [10, p. 13] was developed within six research projects, with the aim of creating a database and digital tools for the analysis of texts in Romanian. The project resulted in the linguistic package "Ortograf", the "Romanian Morphological Dictionary" with over 80,000 words and 2 million forms of inflection and the automatic development for morphological, syntagmatic and phonological analyses [10].

RoWordNet (2001–2013) [10, p. 15] represents the Romanian implementation of the English wordnet. It was part of a larger project, BalkaNET, which targeted several countries in the European Union. The dictionary contained over 20,000 synonym strings and over 36,000 lemmas.

The RoMorphoDict Lexicon (2002–2005) [10, p. 17] is a dictionary of inflectional forms, which also includes hyphenation. It contains 65,000 words, over 770,000 inflectional forms and over half a million syllabic forms.

The RoLEX lexicon (2018–2022) [10, p. 26] project includes the creation of a bimodal corpus for the Romanian language, the development of technologies for natural language processing, automatic annotation of audio data and automatic speech recognition, text-speech synthesis with expressiveness. The result was a lexicon with over 330,000 forms of inflection.

The method of Solomon Marcus

The method of Solomon Marcus [7, p. 118] focuses on the determination of rhythmic structures and the analysis of their characteristics in any type of text. It computes the following metrics:

- **Rhythmic structure** – starts from the distance between two stressed syllables. The rhythmic structure uses the list of accents corresponding to the structure. It counts how many unstressed syllables are until a stressed syllable is encountered. For example, for the list of accents "1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1", the following rhythmic structure "1, 4, 2, 2, 2, 1" is obtained.
- **Rhythmic length** - the number of stressed syllables in a phrase, which is also the length of the rhythmic structure.
- **Rhythmic index** - mathematical ratio calculated based on the rhythmic length and the length of the phrase.
- **Lower/upper rhythmic margin** - the lower/upper dimension of rhythmic structures.
- **Rhythmic diameter** - the difference between the upper rhythmic border and the lower rhythmic border.

The method of Mihai Dinu

The method of Mihai Dinu [8, p. 162] is focused on poetry and determines the rhythmic units and evaluates their features, computing:

- **Monostressed rhythmic unit** - set consisting of one stressed syllable and one or more unstressed syllables.
- **Rhythmic unit length** - the number of syllables in a rhythmic unit.
- **Rhythmic unit rank** - the position of the stressed syllable within a rhythmic unit.
- **Line measure** - the sum of the lengths of the rhythmic units in a line.
- **The number of rhythmic units** in a verse.

The method of Vasile Vasile

The method of Vasile Vasile [9, p. 62] analyzes musicality and repetitive rhythm, focusing on groups of three vowels. Their repetition contributes to defining the rhythm, emphasizing the musicality of the text. They can appear in various combinations, intensifying the sonic impact of the verse. The author divides vowel groups into two types: those containing predominantly open/semi-open vowels, such as "a", "ă", "e", "o" and those containing closed vowels, such as be "u", "â", "î", "i".

IMPLEMENTATION

The language and libraries used

The implementation of the project was done in the Python programming language. Three main libraries used in the development of this project: *tkinter* was used for the development of the graphical interface [11] and *spaCy* for

advanced natural language processing, which uses neural network models for word tagging, text classification, and many other tasks [12]. In this research it was used to extract words and determine their part of speech, being useful in determining stress. The third main library used was *pyphen* for splitting words into syllables using Hunspell dictionaries [13] to hyphenate words that were not in the RoLEX lexicon.

The graphic interface

The project uses a minimalist graphic interface (see Figure 1), easy to use, as it offers several functionalities for analyzing the rhythm of Romanian texts. The GUI features two text boxes. The text input box allows you to write the input text. The output text box displays the result obtained after processing the input.

The functionalities of the buttons in the interface are:

- **Load Input File** - allows the user to load a text file of type txt.
- **Remove Input File** - if an input file has been loaded, this button removes the reference to it, allowing us to use the input box.
- **vers/frază** - select the structure on which the text analysis will be done.
- **Solomon Marcus** - processes the input text using the Solomon Marcus method and displays the result in the output box or output file.
- **Mihai Dinu** - processes input text using Mihai Dinu's method and displays the result in the output box or in the output file.
- **Vasile Vasile** - processes input text using Vasile Vasile's method and displays the result in the output box or in the output file.

Input text and output text display

The input text must be in Romanian. The developed project supports diacritics. Punctuation marks, symbols and numbers are not considered in the analysis of the text according to the three mentioned methods.

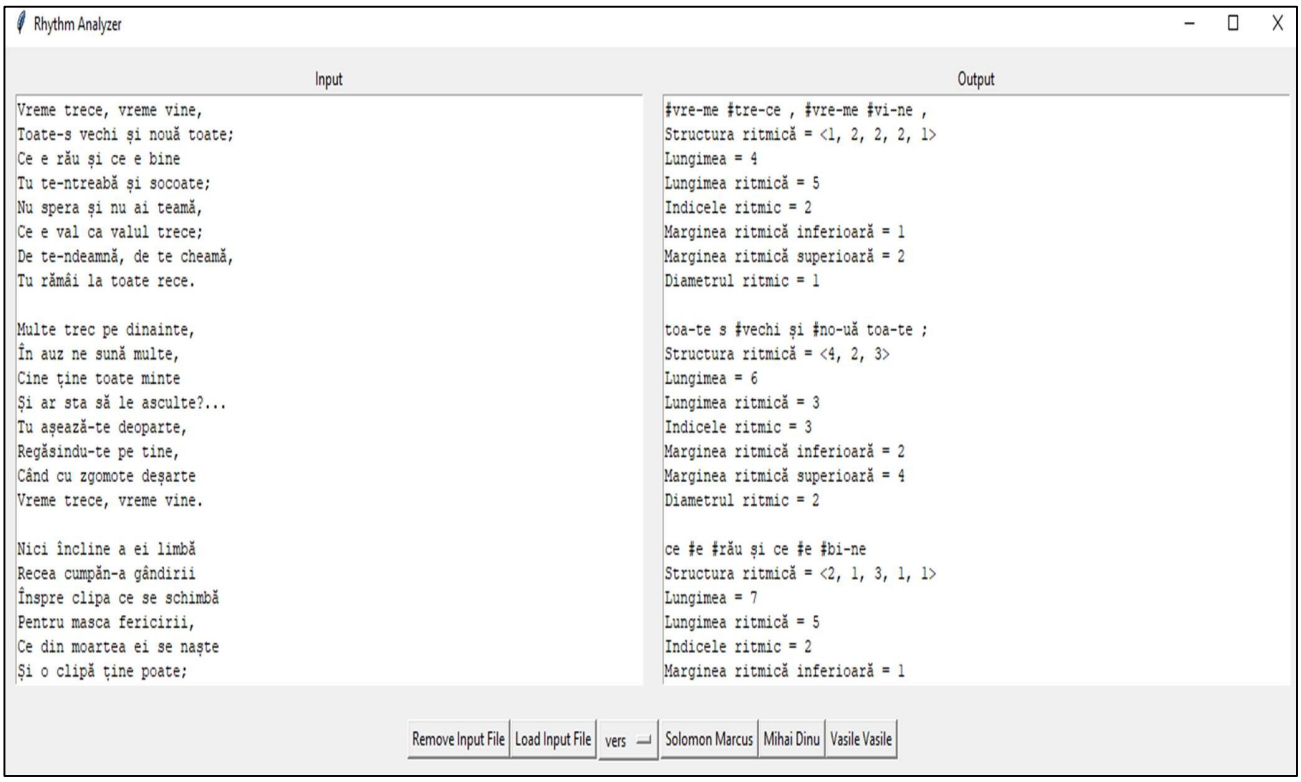

Figure 1. The graphical interface of the project

The results according to Solomon Marcus' method of the verse “Ce e val ca valul trece” are presented as:

ce #e #val ca #va-lul #tre-ce ; Rhythmic structure = <2, 1, 2, 2, 1>
Length = 6 Rhythmic length = 5 Rhythmic index = 2
Lower rhythmic margin = 1 Upper rhythmic margin = 2 Rhythmic diameter = 1

The structure is divided into syllables and the stressed syllable is represented by the "#" symbol, positioned before it. After displaying all the structures and their analyses, the frequency of occurrences for length, rhythmic length, rhythmic index, lower rhythmic margin, upper rhythmic margin, and rhythmic diameter are shown.

The results according to Mihai Dinu's method are:

ce #e | #val | ca #va-lul | #tre-ce; Length_1 = 2 Rank_1 = 2
Length_2 = 1 Rank_2 = 1
Length_3 = 3 Rank_3 = 2
Length_4 = 2 Rank_4 = 1
Distance_1_2 = 1
Distance_2_3 = 2
Distance_3_4 = 2 Predominant binary rhythm. Measure = 8
Acatalectic verse.

The structure is divided into syllables and the stressed syllable is represented by the "#" symbol, positioned before it. The rhythmic units are delimited by the "|" symbol. After all the structures and their analysis are displayed, the frequency of occurrences for length, rank, distance, rhythm, meter, and verse type are displayed.

Displaying results according to Vasile Vasile's method looks like:

ce e val ca valul trece; e e - a | a - a - u Frequency groups:
aee : 50.0%
aau : 50.0%

The structure is displayed together with groups of three vowels identified and delimited by the symbol “|”. After displaying all the structures and their analyses, the frequency of occurrences for the vowel clusters and the frequency of the cluster types are displayed. The order of the vowels is not taken into account in determining the frequency.

Figure 2 presents the structure of the implemented application.

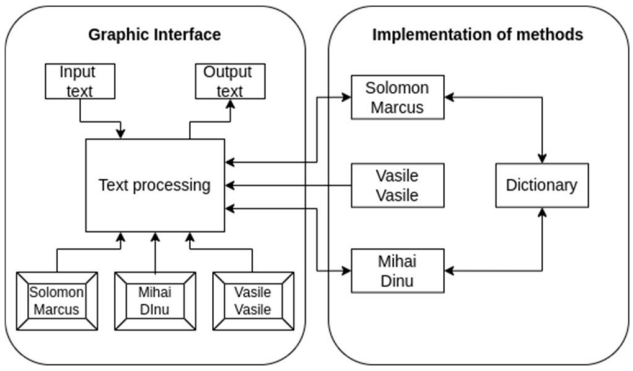

Figure. 2. Project structure

Implementation of functionalities

Implementation of the method of Solomon Marcus

In the implementation of Solomon Marcus' method, the text is divided according to the selected structure:

- verse - the text is divided according to the "\n" character (new line).
- phrase - the text is divided according to the punctuation marks found at the end of a phrase, such as ".", "!", "?".

Afterwards, the following steps are applied for each individual structure:

1. The spaCy library is used to split the structure into words and determine the part of speech of each word. Numbers, symbols and punctuation are not processed.
2. For each word, it is checked to see if it exists. If so, syllable division and stress are determined, considering the form and part of speech. Otherwise, the pyphen library is used for hyphenation. If there are at least two syllables, the stress is placed on the penultimate syllable (After an individual analysis, we noticed that the stress on the penultimate syllable predominates). For a syllable, it becomes stressed.
3. The broken syllables are added to a list containing all the syllables in the structure.
4. The stress for each syllable is added to a list containing all the stresses in the structure. If the word is an adjunct, then it is considered that it does not have an accent (Solomon Marcus does not specify exactly which are the words whose accent is not taken into account, but after an analysis it was noticed that they are the following: conjunctions, pronouns, adpositions).
5. The number of words in a structure is counted in a list.
6. Having all the necessary information, the mathematical formulas can be applied to determine the metrics presented in the previous section rhythm for the analysis on the given text.

After calculating the information for each structure, the language is also analyzed:

- **the rhythmic index of the language** - the maximum value of the indices calculated for each structure.
- **the lower edge of the language** - the minimum value of the lower rhythmic edges.
- **the upper edge of the language** - the maximum value of the upper rhythmic edges.
- **the rhythmic diameter of the language** - the maximum value of the diameters.
- **rhythmic dimension of language** - the difference between the upper edge of the language and the lower edge of the language.

There are also calculated the frequencies of rhythmic length, rhythmic lower edge, rhythmic upper edge, rhythmic diameter and rhythmic index using frequency dictionaries.

Implementation of the method of Mihai Dinu

In the implementation of Mihai Dinu's method, almost the same methods of text analysis are used as in Solomon Marcus's method, thus obtaining information about the division of words into syllables, stressed syllables and the length of the structure.

In addition, Mihai Dinu's method also uses a list for each structure, which provides information about the type of word: joining word or stressed word. For example, for the structure "ce e val, ca valul trece", we get the list "0, 1, 1, 0, 1, 1".

The rhythmic units are created taking into account the list presented above. A rhythmic unit comprises joining words and a single stressed word. Thus, for the example "ce e val, ca valul trece" which has the following list with the type of words "0, 1, 1, 0, 1, 1", we obtain the rhythmic units "ce e", "val", "ca valul", "trece".

The mathematical formulas are applied to determine the rhythm analysis on the respective text, using Mihai Dinu's technique, as described in the previous section. Also the following metrics are computed:

- **The distance between consecutive rhythmic units** - represents the distance between two accented syllables.
- **The type of rhythm** - is determined taking into account the distance.
- **Measure** - represents the number of syllables in a structure.
- **Type of structure** - it is calculated according to the measure and type of rhythm.

For this method are also computed the frequencies of rhythmic unit length, rhythmic unit rank, distances between rhythmic units, rhythm types, measures, and structure types using frequency dictionaries.

Implementation of the method of Vasile Vasile

The same algorithm as in the previous methods is used to divide the text into structures according to the chosen type, verse or phrase. Each structure is traversed and consecutive groups of as many vowels are created.

Frequency dictionaries are created for three-vowel sets and for the vowel group type. A cluster containing at least two open or semi-open vowels is considered an open cluster. A group containing at least two closed vowels is considered a closed group.

RESULTS

The implemented methods were used for analyzing four genres of texts: lyric, prose, speech, and scientific.

Results obtained on lyric text

An analysis is presented carried out on the poem "Glossa" by Mihai Eminescu [15], analyzing the rhythm according to the methods described by Solomon Marcus, Mihai Dinu, Vasile Vasile.

The method of Solomon Marcus - poetry

The structure according to which the text is divided is the verse. It is noted that the lyrics are between 4 and 6 words long. There is an almost equal distribution between them. The rhythmic length is predominantly 3. The rhythmic index is 2 in more than 80% proportion. This indicates a regular pattern of two rhythmic units. The lower rhythmic edge is 1. The predominant upper rhythmic edge is 4, with significant values of 3, 2. The rhythmic diameter includes the values 3, 2, 1, with the highest frequency being 3. The language has the following characteristics: rhythmic index - 4, lower rhythmic border - 1, upper rhythmic border - 8, rhythmic diameter - 7, rhythmic dimension - 7.

The method of Mihai Dinu - poetry

It is observed that in the given poem the rhythmic units having the length of 2, 3 and 4 syllables predominate. The most common rank is 1, 2 and 3. The distance, representing the distance between two stressed syllables, is 2, a fact that is also observed by the dominant type of rhythm, the binary rhythm, with a percentage of 52%. The majority measure is
8. Acatalectic verses are found in a large proportion, emphasizing the musicality through the balance formed between the measure of the verses and the type of rhythm.

The method of Vasile Vasile - poetry

It can be observed that the "eee" group has a fairly high percentage, although it is not a group of vowels often encountered in the Romanian language. This is due to the repetition used in the given text, "time passes, time comes", which has the role of emphasizing the passage of time, creating a parallel between the past and the future. Also, the text presents a frequency of 71.3% for open/semi-open vowel groups and a percentage of 28.7% for closed vowel groups.

Results obtained on prose

The first chapter of the literary work "Baltagul" by Mihail Sadoveanu [16] is analyzed. The text includes both long sentences with archaic language and dialogue sequences.

The method of Solomon Marcus - prose

It is observed that phrases with lengths of 8, 7, 6 are the most common. The dominant rhythmic length shows values 7, 6 and 9, all with almost the same percentage of occurrence. The majority rhythmic index is 2, with a very high percentage of 86%. This fact indicates the presence of binary rhythm. The dominant lower rhythmic edge is 1, and the upper rhythmic edge is 4. Thus, the predominant rhythmic diameter of 3 is noted. Also, the presence of an atypical dynamic rhythm, having a value equal to 0. It indicates the presence of phrases of reduced size, made up only from linking words, which are not stressed. The language has the following characteristics: rhythmic index - 5, lower rhythmic edge - 1, upper rhythmic edge - 10, rhythmic diameter - 9, rhythmic dimension - 9.

The method of Mihai Dinu - prose

The dominant length of rhythmic units is 3. The majority rank is 2 and 1, both values having a similar percentage. The predominant distances are 2 and 3, with values of 30% and 24%. The rhythm is binary, in most of the text. However, a ternary rhythm is also identified for 30% of the text. The measure shows various values, not having a percentage that differentiates significantly. The acatalectic phrase is present in 62%.

The method of Vasile Vasile - prose

The group of vowels "aei" has a dominant percentage compared to the other groups of vowels. Also, open/semi-open vowel groups are present with a frequency of 66%, compared to closed ones, 34%.

Results obtained on speech

A political speech ("Sedinta comuna a Camerei Deputatilor si Senatului din 21 decembrie 2014" [17]) is analyzed according to the methods presented by Solomon Marcus, Mihai Dinu and Vasile Vasile.

The method of Solomon Marcus - speech

The phrases used in this speech present a large number of words, the predominant lengths being 20, 17, 10. The rhythmic length presents mostly higher values, 7, 10, 14, than in the case of the analyzed poetry and prose. The rhythmic index, with a frequency of 92%, is 2, emphasizing the binary rhythmicity of the text. The lower rhythmic border is 1, in most of the text. The upper rhythmic border includes the following dominant values, 6, 5, 7, the value 6 having a percentage of 34%. The majority diameter is 5. The language has the following characteristics: rhythmic index - 3, lower rhythmic edge - 1, upper rhythmic edge - 10, rhythmic diameter - 9, rhythmic dimension - 9.

The method of Mihai Dinu - speech

The majority of rhythmic unit lengths are 3, 2 and 4. The rank is predominantly 1 and 2, with a percentage of 29% and 27% respectively. The dominant distance between two stressed syllables is 3. The presence of the ternary rhythm in 45% and the binary rhythm in 42% is noted. It can be seen that the phrases have a fairly high measure value, 19, 34 and 39. The phrases are acatalectic in proportion to 60% and catalectic 40%.

The method of Vasile Vasile - speech

As in the case of the prose text, the group of vowels "aei" has the highest percentage. Also, the proportions of groups of open vowels 66% and closed vowels 34% are preserved.

Results obtained on scientific text

A scientific text about animals ("Regnul Animalia", from Wikipedia [18]), which contains technical terms and names in Latin, is analyzed. Even if rhythm is not an important element in scientific texts, some of its characteristics are observed in the following analyses.

The *Method of Solomon Marcus - scientific text*

The dominant lengths are 13, 29, 24 and 19. The most frequently encountered rhythmic length is 19. Rhythmic index 2 is present in over 97%. The dominant lower rhythmic edge is 1. The majority values for the upper rhythmic edge are 5 and 6. The language has the following characteristics: rhythmic index - 3, lower rhythmic edge - 1, upper rhythmic edge - 9, rhythmic diameter - 8, rhythmic dimension - 8.

The Method of Mihai Dinu - scientific text

The length of the rhythmic unit is predominantly 3. The majority rank is 2. The distance between two stressed syllables is 3, 4 and 2, the dominant being the first value with a frequency of 28%. The majority rhythm is the ternary one. There are phrases with a large size, having the measure of 42, 65, 45, 71 with a percentage of 5%. The phrases are mostly acatalectic.

The Method of Vasile Vasile - scientific text

The group of vowels "aei" has a dominant percentage, 11%, compared to the other groups of vowels. Also, open/semi- open vowel groups are present with a frequency of 72%, compared to closed ones, 28%.

CONCLUSIONS

The research presented herein aimed to automate the analysis of the rhythm of Romanian texts, using the methods offered by Solomon Marcus, Mihai Dinu and Vasile Vasile.

Solomon Marcus approaches a mathematical analysis of the rhythm used in Romanian texts. The rhythmic structure is created, taking into account the stressed syllables within a phrase. Based on this structure, the rhythmic index, rhythmic length, lower rhythmic edge, upper rhythmic edge are determined. Also, by evaluating the entire text, the rhythmic index of the language, the diameter of the language, the rhythmic dimension is also determined.

Mihai Dinu offers a similar approach to that of Solomon Marcus, but he divides the verses into rhythmic units, which contain a stressed word and unstressed words. Taking into account the rhythmic units, the rank, diameter, length, rhythm type, meter and verse type are determined.

Vasile Vasile analyzes groups of three vowels, taking into account their type: open, semi-open and closed.

For the application of the first two methods, it was necessary to divide words into syllables and determine the stressed syllable. This was possible with the help of the RoLEX dictionary, which provides information on the morphosyntactic structure, hyphenation, stressed vowel and pronunciation for more than 330,000 words. Using the spaCy library, the part of speech of a word is decided by doing a dictionary search for the word and its part of speech. If it was not found in the dictionary, the *pyphen* library was used for dividing into syllables and the penultimate syllable was stressed, if there were at least two. Thus, the text was ready to be analyzed according to the desired method.

The results obtained on the examples provided by the three authors are similar to their analyses. In the case of the methods of Solomon Marcus and Mihai Dinu, there was a change generated by the different syllable division of some words. In the case of Vasile Vasile's method, several groups of vowels were selected differently. This fact happened because the author does not specify how the vowels should be selected.

Four types of texts were analyzed: poetry, prose, speech and scientific text. The features of rhythm were very well highlighted in the case of poetry. The accented repetition of some groups of vowels was also identified. In the case of the other types of text, where rhythm is not a main element, certain characteristics of it have been identified.

In conclusion, the developed application offers an automatic way of rhythm analysis, managing to evaluate texts of a considerable size in a relatively short time, providing conclusive data about its characteristics.

REFERENCES

1. Niculescu, I.D., Trausan-Matu, S. Rhythm analysis in chats using Natural Language Processing, *Romanian Conference on Human-Computer Interaction (RoCHI)*, (2016), pp. 107-112.
2. Rhythm. [Online]. Available: https://dexonline.ro/intrare/ritm/48984 [accessed at 12.06.2024].
3. Balint, M., Dascalu, M., & Trausan-Matu, S. The Rhetorical Nature of Rhythm. In *15th Int. Conf. on Networking in Education and Research (RoEduNet)*. Bucharest, Romania: IEEE (2016), pp. 48–53.
4. Balint, M., Dascalu, M., & Trausan-Matu, S. Classifying Written Texts through Rhythmic Features. In *17th Int. Conf. on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA 2016)*, Varna, Bulgaria: Springer, (2016), pp. 121–129.
5. Accent. [Online]. Available: https://dexonline.ro/definitie/accent [accessed at 12.06.2024].
6. Lőrincz B, Irimia E, Stan A, Barbu Mititelu V. RoLEX: The development of an extended Romanian lexical dataset and its evaluation at predicting concurrent lexical information. *Natural Language Engineering* Cambridge University Press, 2023;29(3), doi:10.1017/S1351324922000419, (2022), pp. 720-745.
7. Marcus, S. *Poetica Matematică*: Editura Academiei, (1970).
8. Dinu, M. *Ritm și rimă în poezia românească*. Editura Cartea Românească, (1986).
9. Vasile V. *De la muzica firii și a sufletului la muzica sferelor,* Editura Petrion, (2005).
10. Teodorescu, C.-M. Un inventar al lexicoanelor (digitale) pentru limba română. Perioada 1973-2022, *Anuar de lingvistică și istorie literară*, (2023).
11. tkinter — Python interface to Tcl/Tk, 2007. [Online]. Available: https://docs.python.org/3/library/tkinter.html [accessed at 16.06.2024]
12. spaCy 3.7.5, 2024. [Online]. Available: https://pypi.org/project/spacy/[accessed at 16.06.2024]
13. pyphen 0.15.0, 2024. [Online]. Available: https://pypi.org/project/pyphen/ [accessed at 16.06.2024]
14. Tașcu, V. Teoria ritmurilor în artă și poezie. *Dacoromania. Serie nouă*, XI-XII, p. 207-234 Editura Academiei, (2007).
15. Eminescu, M. Glossă, 2024. [Online]. Available: http://www.romanianvoice.com/poezii/poezii/glossa.php
16. Sadoveanu, M. *Baltagul*, Editura Mihail Sadoveanu, (2024).
17. Discurs Klaus Iohannis ceremonie de depunere a jurământului de învestitură în funcție, 2024. [Online]. Available: https://www.presidency.ro/ro/media/discursuri/discursul-sustinut-de-presedintelui-romaniei-domnul-klaus-iohannis-in-cadrul-ceremoniei-de-depunere-a-juramantului-de-investitura-in-functie
18. Regnum Animalia, 2024. [Online]. Available: https://ro.wikipedia.org/wiki/Regnul_Animalia