

# A Program For Analyzing Classical Arabic Poetry For Teaching Purposes

Munef Abdullah Ahmed<sup>1</sup>, Stefan Trausan-Matu<sup>1,2,3</sup>

<sup>1</sup>Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Romania

<sup>2</sup>Research Institute for Artificial Intelligence of the Romanian Academy, Bucharest, Romania

<sup>3</sup>Academy of Romanian Scientists, Bucharest, Romania

E-mail: [yabrahmun@yahoo.com](mailto:yabrahmun@yahoo.com), [trausan@gmail.com](mailto:trausan@gmail.com)

**Abstract.** A difficult but important aspect of e-learning for English language speakers is teaching the Arabic language. The major task for students in this context is analyzing Arabic poetry, which differs in several aspects from the English poetry. Natural Language Processing (NLP) for poetry analysis, and especially, being excellent in finding Arabic poetic characteristics is extremely difficult. Despite the rapid progress in this area in some international languages, the analysis of classical Arabic poetry has not received a sufficient attention due to the difficulty of the Arabic language, and the difficulties of analyzing its poetic theories. In this paper, a new way of finding the characteristics of the classical Arabic poem was introduced. One of the most important characteristics of Arabic poetry is the rhythm which is considered as the crown of this type of poem. The rhyme, which is the end of the musical tone of each verse in a poem, and the types of rhyme are also important features of any poetic art. In this article, we focused on the short vowels (“dammah”, “kasra”, “fatha”, “tanween”, and “shade”), mobile and constant letters which depend on the theory that analyzes the characteristics of the classical Arabic poetry introduced by Al-Khalil bin Ahmed Al Farahidi. An analysis program was written in a Python platform, which analyzes poems and provides the aforementioned features and some statistical measures. The program will be developed for an e-learning platform, and for use in the classification and identification of classical Arabic poems

**Keywords:** e-learning; Natural Language Processing; Arabic Language; Poetry analysis; Rhythm.

## 1. Introduction

Natural Language Processing (NLP) is a domain of artificial intelligence which is closely linked with linguistics. We can see the intertwining between the two sciences, linguistics ideas being used in computer science techniques for analysis. NLP helps in the production and improvement of software for analysis of texts in natural language. It also stimulates the understanding of these languages. The most used languages (English or French) have attracted much research attention and there has been an important progress in developing NLP applications. The Arabic language has failed to attract such enormous attention and for this reason, this paper introduced the use of NLP for the study and development of a program for the teaching and analysis of classical Arabic poetry. Classical Arabic poems were chosen because they are the source of most actual rules in Arabic poetry, as well as being the foundation for Arabic poetry (Ahmed & Trausan-Matu, 2017; Kaplan & Blei, 2007).

In this paper, we start with the state of art, and then, with the general features of this field of study. The five types of Arabic poetry and their features are also discussed. The rhythm is a basic characteristic of these classes which is used as a feature to distinguish them. The short vowels are very important because any change in them will also result in a change in their meaning. They are also considered as a distinguishing feature in the Arabic language. In the “Al Arud” section, the moving and constant letters were classified and discussed. The “Tafilah” from the main part of the “Al Arud” contains ten types and is divided into two types based on the number of letters. The “Al Buhur”, which depends on the “Tafilah”, is of sixteen kinds, as introduced by Al-Khalil bin Ahmed Al Farahidi. The rhyme for this class of poems is also an important feature. The rhyme is the musical ending of the poetic verses, a condition for classical Arabic poems, and is divided into six types. The automatic calculation of the characteristics of the classical Arabic poem is presented as well. The study finally ended with a conclusion of the work.

## 2. State of the art

One of the most important natural languages in the world is the English language. It is also one of the first that used NLP for finding the characteristics of the language, and one of the first languages to use NLP and

mathematics formalisms in the poetic analysis. Preliminary studies in this field analyzed American poems and performed clusterings based on the style of the poem. The first studies for the Arabic language classified the Arabic text (Nehar, Ziadi, Cherroun, & Guellouma, 2012) and performed poem recognition (Tizhoosh & Dara, 2006). The text classification technique has been used by some researchers; this technique used in NLP for distinguishing between poems and general text in the Arabic language depends on some important characteristics such as rhyme, meter, and rhythm (Tizhoosh, Sahba, & Dara, 2008). The theory of Al-Khalil bin Ahmed has been used by researchers for analyzing Arabic poetry meter both to written and spoken Arabic poems (Saleh & Elshafei, 2012).

The types of poetry in the Arabic language vary according to their form. The following types of Arabic poetry can be distinguished: “Al Hur”, “Al Rubaiat”, “Al Mursel”, “Al Nathur”, and classical Arabic poetry (Jayyusi, 1977). In addition to the several features used for classifying poetry, such as syntactic, phonemic, orthographic, and lexical features, poems are divided or classified based on the way they are written by novices or experts (Saleh & Elshafei, 2012).

Repetition is also an important feature that is being studied by several researchers (Trausan-Matu, 2012) and is considered in this paper as well. A research on Malay poetry also classified it into different themes (Jamal, Mohd, & Noah, 2012). Finding and recognizing classical Arabic poems depend on several features such as the rhyme, words, style, and structure (Almuhareb, Alkharashi, Saud, & Altuwaijri, 2013). Machine learning has been used in the classification of the emotion in Arabic poetry into Fakher, Retha, Ghazal, and Heja (Alsharif, Alshamaa, & Ghneim, 2013). Arabic texts have also been classified using polynomial neural networks (Al-Tahrawi & Al-Khatib, 2015). Based on several features such as repetition, punctuation, rhyme, and text alignment, the modern Arabic poems can be recognized from the other non-poem texts (Almuhareb, Almutairi, Al-Tuwaijri, Almubarak, & Khan, 2015).

There are several other features used in the study of poetry (Rakshit, Ghosh, Bhattacharyya, & Haffari, 2015). Phonemic, as one of these features (Fernandez, 2004) is related to the sound devices used in the poem (Balint, Dascalu, & Trausan-Matu, 2016b). The second important feature in Arabic poetry is the type of the letters (mobile or quiescent) according to Al-Khalil bin Ahmed’s theory (Drory, 2000). The structure of the sentences and the

frequency of the parts of speech are important characteristics commonly referred to as syntactic features (Martin & Jurafsky, 2009). Other features that deal with the words and the relation between words, called lexical features (Lefer, 2011) are the orthographic features, the number of words, lines, sentences, and their average (Balint, Dascalu, & Trausan-Matu, 2016a; Niculescu & Trausan-Matu, 2016).

### **3. Classical Arabic poetry**

Classical Arabic poetry is considered as the origin of all Arabic poetry, from which other forms of Arabic poetry were derived. The poems in the classical Arabic poetry contain multiverses depending on the author and the purpose of the poem. Each verse consists of two parts of equal length; the first part is called “sadr” and the second part is called “ajuz”. The form of classical Arabic poem is of three types (Almuhareb et al., 2013). The term “buhur”, which was invented by Al-Khalil bin Ahmed Al Farahidi, is used for metering the rhythmic system in a poem, and the measurement unit for “buhur” is known as “Tafilah”. In this type of poetry, the addition or removal of any letter from any verse will change the meter and the number of “Tafilah” in the verse. Therefore, this feature helps critics and scholars to identify the quality of an Arabic poem. Another important feature of the classical Arabic poetry is the rhyme “qufiah”. In general, all the verses in the poem have the same rhyme and these verses are described based on their number, as follows: one verse is called “Yetim”, two and three verses are called “Natka”, four, five, and six verses are called “Kuteah”, while seven or more verses are called “Kassed” (van Gelder, 2012).

### **4. Classical Arabic poem rhythm**

The rhythm of a natural phenomenon means the regular sequence of a group of elements. Clock beats, heartbeats, dancers' steps, poets' poetry, and even olfactory scents, symmetrical drawings of painters and engineers, eyesight, tactile sensations, and taste solvents are all characterized by regular rhythms. In some cases, the rhythms rise to a higher degree, and in the most complex case, they are varied in a beautiful consistency to form different fine arts, including poetry. The rhythm is a necessity of modern life in general, and it is evident in the human expressions when speaking. These expressions may

illustrate the tension, psychological state or feeling, and this rhythm is understood by humans. Poetic rhythm is a fundamental property that was not imposed from an external source. The organization of the voices of a language to follow a specific time pattern is no doubt an organization of the language sounds. In the classical Arabic poem, Al-Khalil bin Ahmed Al Farahidi defined the rhythm, saying that “equilateral movements have successive returns”. He called the science used in the study of the Arabic poetry “Al Arud” science. This definition was used in this study to find the rhythm in classical Arabic poems (Ahmed & Trausan-Matu, 2017).

#### 4.1 Short vowels in classical Arabic poem

The classical poem is based on Arabic language and it masters in this field. This language is considered to be complex and contains many characteristics and ramifications. One of the characteristics are the short vowels, which are placed above or under the letter in the words. Short vowels are very important in the Arabic language; any change in the type of the short vowels changes the meaning and pronunciation. (Ahmed & Trausan-Matu, 2017).

Table 1: The short vowels with “shadde” and “tanween” in the classical Arabic poem

The short vowel	The Sign	Applied to the letters	Pronunciation
“sukoon”	◌ْ	كَمْ - كُ	M - K
“dammah”	◌ُ	كُم - كُو	Mu - Ku
“Kasra”	◌ِ	كِم - كِي	Mi - Ki
“Fatha”	◌َ	كَم - كَا	Ma - Ka
“tanween fatha”	◌ً	كُمَّ - كَأَ	Man - Kan
“tanween kasra”	◌ًِ	كِمِ - كِيِ	Min - Kin
“tanweendamhah”	◌ًُ	كُمُ - كُوُ	Mon - Kon
“shadde”	◌ّ	كَمّ - كَاّ	Mm - Kk

#### 4.2 Classical Arabic poem “Al Arud” section

In the Arabic language science, the term “Al Arud” describes the smallest part of speech that can be pronounced separately from the others. It consists of at least two characters and may be increased to a maximum of five characters in the “Al Arud” section. In the Arabic language, it is not possible to start with a constant letter. The scholars divide the “Tafilah” into different

sections with different movements in their number and letters, as illustrated in the following “Al Arud” section (van Gelder, 2012):

- “Al Sabeб” is a music section, consisting of two letters, and divided into two types:
  - “Sabeб Kafif” - This music section begins with a moving letter, followed by a constant letter (/0). The examples of this section are the Arabic words “مَنْ”, “كَيْ”, and “لَمْ”.
  - “Sabeб Thkil” - In this music section, the first and second letters are moving letters (/). The examples for this section are the Arabic words “لَمْ” and “لَكَ”.
- “Al Awtad” - This term refers to the music section that consists of three letters. There are two types of this music section:
  - “Wtad majmoh” - This music section contains two moving letters, followed by a constant letter (/0). The examples of this type are the Arabic words “لَقَدْ”, “مَنْيَ”, and “أَنْهَ”.
  - “Wated mafroq” - This music section consists of two moving letters, with a constant letter between them (/0/). The examples of this section are “أَنْتَ”, “فَالِ”, and “صَامَ”.

### 4.3. Classical Arabic poem “tafila”

This is an important term in the classical Arabic poems. In the Arabic poetry, “Al Buhur” depends on the “Tafila”. This term consists of more than one “Al Arud” section which may be one “Al Sabeб” and one “Al Awtad” or two “Al Sabeб” and one “Al Awtad”. It is classified into two types based on the number of the letters. The first type consists of five letters which include “Wtad majmoh” and “Sabeб Kafif”. Table 2 illustrates this case.

Table 2: The five letters of “Tafilah”

Name of “Tafilah”	“Al Arud” writing	Arabic Word
فَعُولُن (Faulun)	//0/0	أَلَيْنَا
فَاعِلُن (Failun)	/0//0	هَذَا أَتَى

The second type consists of seven letters which include two “Al Sabeб” and one “Al Awtad”. It has three types; the first type is the “Tafilah” which contains “Sabeб Kafif” as illustrated in Table 3.

Table 3: The first type of the “Tafilah” consisting of seven letters

Name of “Tafilah”	“Al Arud” writing	Arabic word
مفاعلتن (Mufalatun)	//0//0	فخذبيدي
متفاعلتن (Mutafilun)	///0//0	لكأخرفي

The second type is the “Tafilah” which contains “Wtad majmoh” and two “Sabebe Kafif” as illustrated in Table 4.

Table 4: The second type of the “Tafilah” consisting of seven letters

Name of “Tafilah”	“Al Arud ” writing	Arabic word
مفاعيلن (Mafailun)	//0/0/0	أناديكم
مستفعلن (Mustafilun)	/0/0//0	لأتحزنو
فاعلاتن (Failatun)	/0//0/0	ياصديق

The third type is the “Tafilah” which contains “Wated mafroq” as illustrated in Table 5.

Table 5: The second type of the “Tafilah” consisting of seven letters

Name of “Tafilah”	“Al Arud” writing	Arabic word
مفعولات (Mafulat)	/0/0/0/	مات شمرز
فاعلاتن (Failatun)	/0//0/0	نام صحنبي
مستفعلن (Mustafilun)	/0/0//0	باطيرقم

The difference between these types of “Tafilah” comes from the differences in the arrangement of the “Al Arud” sections, which depends on the position of the constant and moving letters. The classical Arabic “Al Buhur” poem has sixteen kinds, and each of these is made up from “Tafilah”, maybe by repeating the same “Tafilah” or by combining several “Tafilah”. Using this method of “Al Buhur” writing for the classical Arabic poem, it is possible to determine whether a poem matches the rules of classical Arabic poems or not.

#### 4.4. Classical Arabic poem rhyme

In the Arabic poetry, there are several definitions of rhyme- a specific part that exists in the last part of each verse. Some researchers gave this name for

the last word in each verse of a poem. Some scholars suggest that the rhyme is the name for the last “Tafilah” in each verse, while others gave referred to rhyme as the last letter in every verse of a poem (Almuhareb et al., 2013).

The most famous researcher in the field of Arabic poetry Al-Khalil bin Ahmed Al Farahidi gave the most important definition of rhyme. He started from the end of each verse in a poem to the beginning of the verse, and the stationary letter “harf sakin”. They rhyme may be one or more letters between the end of a verse and the first moving letter “harf mutaharrik”. Hence, the rhyme may be part of a word, a word or two words. This definition is used in this paper.

## 5. Methodology

In our works, the well-known spelling rules are not enough and this requires the adoption of the following rules: letters with pronunciation are written while letters without pronunciation are not written. Therefore, some letters must be added:

- If there are “shadde” above the letter, the letter must be doubled to two letters. The first is static and the second is moving. For example, the Arabic word “رَق” must be written “رُقُق”.
- If there is any type of “tanween” above or under the letter, it must be replaced by the letter “ن”. From the application of this rule, the Arabic word “جِبْلٌ” is written “جِبِلن”. The letter “ا” is added to some sign names, therefore, the Arabic word “هؤلاء” must be written “هؤلاءا”.
- The letter “و” is added to the word if it ends with the pronoun “ه” above the sign “dammah. For example, the word “لله” should be written as “لهو”.
- The letter “ي” is added to the word if it ends with the pronoun “ه” under the sign “Kasra”. For example, the word “فيه” should be written as “فيهي”.
- When the short vowels “dammah”, “Kasra” and “Fatha” come at the end of the “sadr” or “ajuz” of the verse, it must change to letter “و”, “ي”, and “ا”.

The flowchart of the program is depicted in Figure 1.

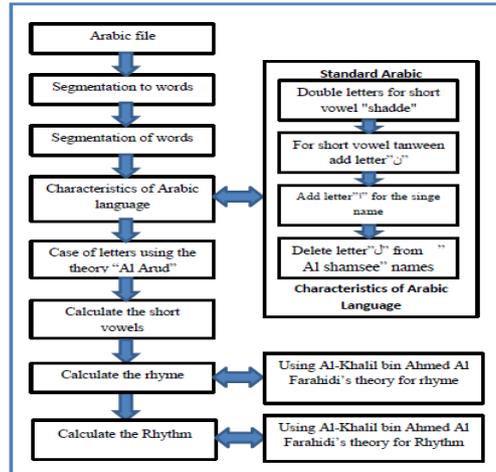


Figure 1. Flow diagram of the program

## 6. Results

A program was implemented for the calculation of several features (rhyme type and percent, the case of letters, and short vowels) of classical Arabic poems. For educational purposes for which the program was developed, both the original, Arabic language and English language versions of the poem were displayed. The results of the proposed method are presented in Figures 2 to 5.

In the beginning, the case of the letters was determined and used in the classical Arabic poems. This is an important fact in our work because the theory "Al Arud" depends on this feature. From the performed analysis, the percentage of the mobile letters in classical Arabic poems was found to be greater than that of the constant letter. The average percentage of the mobile letters was approximately 70% and this is one of the features of the Arabic language. Figure 3 illustrates this percentage in a poem.

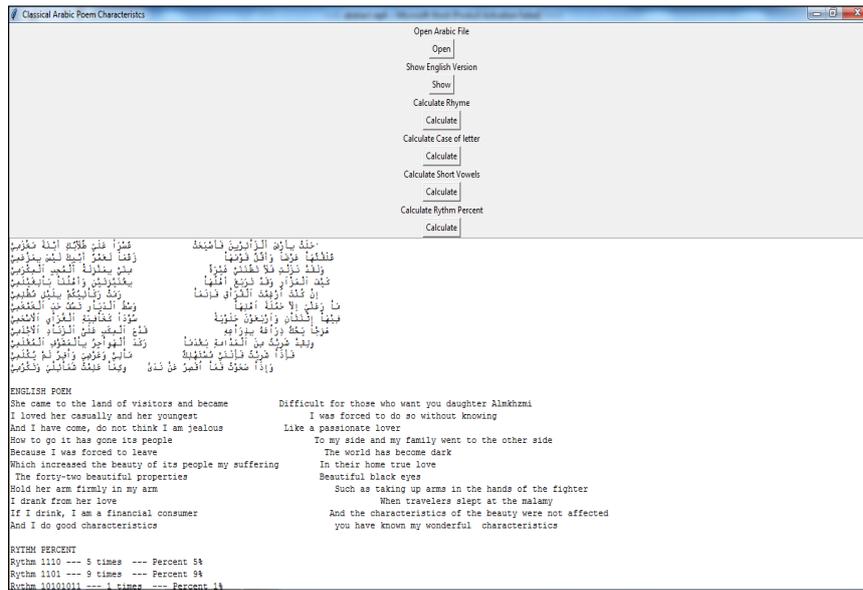


Figure 2. The program interface

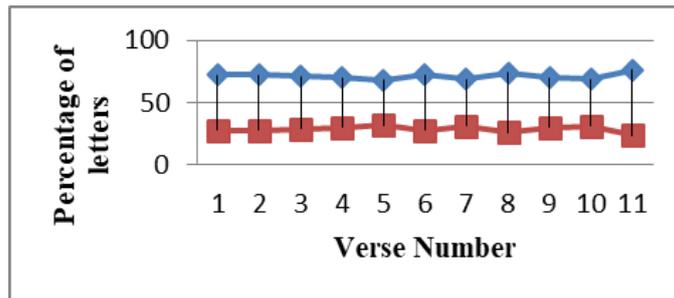


Figure 3. Percentage of the constant and mobile letters in a classical Arabic poem with 11 verses

The short vowels are also an important feature in the Arabic language which has a direct effect on the “Al Arud” theory. There are several kinds of vowel, including “dammah”, “fatha”, “shadde”, “kasra”, and “tanween”. The effect of the short vowels can influence the pronunciation of the letters above or under them. Figure 4 showed the percentage of vowels in each verse. From the figure, a descending trend of vowel influence was observed in the order of “fatha”, “kasra”, “dammah”, “shadde”, and “tanween”. From the analysis, it was observed that 80% of the classical Arabic poems have the same order

of short vowels.

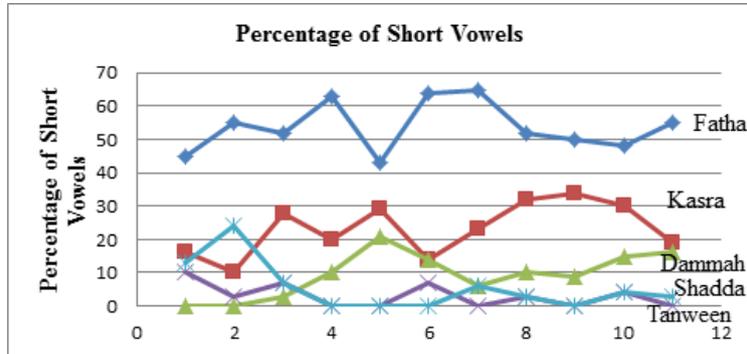


Figure 4. Percentage of short vowels in a classical Arabic poem with 11 verses

We calculated the rhyme for the classical Arabic poem starting from the Al-Khalil bin Ahmed Al Farahidi’s theory. The type of rhyme is of interest in relation to the “Al Roy” letters. The rhyme is an extremely beautiful thing in the classical Arabic poems which gives it a musical sound. Table 6 illustrates the rhyme for a classical Arabic poem with 11 verses. The rhyme type depends on the number of mobile letters between the last two constant letters, and there are five types.

Table 6. Rhyme words and rhyme type for a classical Arabic poem with 11 verses

No. verse	Rhyme Words	Rhyme type
1	مَحْرَمِي	Al Mutadariq
2	بِمَرْعَمِي	Al Mutadariq
3	الْمَكْرَمِي	Al Mutadariq
4	بِالْعَلَمِي	Al Mutadariq
5	مُطَلَبِي	Al Mutadariq
6	الْخَمْرَمِي	Al Mutadariq
7	الْأَمْرَمِي	Al Mutadariq
8	الْأَجْرَمِي	Al Mutadariq
9	الْمَطْرَمِي	Al Mutadariq
10	يَكْرَمِي	Al Mutadariq
11	وَالْكَرَمِي	Al Mutadariq

Rhythm is considered as one of the main features of classical Arabic poems in addition to rhyme because they complete each other. The loss of

either feature in classical Arabic poetry determines the loss of the poetic property. In our works, this feature was calculated by the pronunciation of the words, not only by their writing based on Al-Khalil bin Ahmed Al Farahidi's theory. Figure 5 depicts an example of the cases considered in classical Arabic poems.

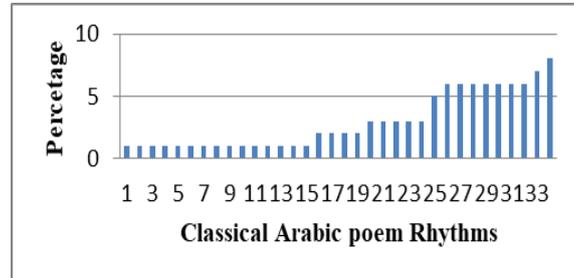


Figure 5: Percentage of the classical Arabic poem rhythm with 11 verses

In this study, the working language was standard Arabic. The normal language cannot be used in calculating the characteristics of the classical Arabic poem because it does not contain movements and short vowels which are the basis for the theory used in this study, and in any work, that deals with classical Arabic poems.

## 7. Conclusion

In this study, we investigated the use of natural language processing as a tool in analyzing the classical Arabic poems. The adopted strategy helped in the implementation of automatic methods for finding the characteristics of the types of Arabic poetry which can be used for teaching purposes. Among these characteristics, the constant and mobile letters were determined, which are the main concepts based on the “Al Aurde” science, as used in our study. We also detected short vowels in the context of “Tafilah” and how the verse will be changed. Rhythm, a major feature of this type of poetry, together with the rhyme and its type, complements the music tune for classical Arabic poems. For the future works, this method will be improved to suit its application to the other types of Arabic poetry. The results obtained from the different applications and characterizations can be used to detect the type and quality of Arabic poetries.

## References

- Ahmed, M. A., & Trausan-Matu, S. (2017). *Using natural language processing for analyzing Arabic poetry rhythm*. Paper presented at the Networking in Education and Research (RoEduNet), 2017 16th RoEduNet Conference.
- Al-Tahrawi, M. M., & Al-Khatib, S. N. (2015). Arabic text classification using Polynomial Networks. *Journal of King Saud University-Computer and Information Sciences*, 27(4), 437-449.
- Almuhareb, A., Alkharashi, I., Saud, L. A., & Altuwaijri, H. (2013). *Recognition of classical Arabic poems*. Paper presented at the Proceedings of the Workshop on Computational Linguistics for Literature.
- Almuhareb, A., Almutairi, W. A., Al-Tuwaijri, H., Almubarak, A., & Khan, M. (2015). Recognition of Modern Arabic Poems. *JSW*, 10(4), 454-464.
- Alsharif, O., Alshamaa, D., & Ghneim, N. (2013). Emotion classification in Arabic poetry using machine learning. *International Journal of Computer Applications*, 65(16).
- Balint, M., Dascalu, M., & Trausan-Matu, S. (2016a). *Classifying Written Texts through Rhythmic Features*. Paper presented at the International Conference on Artificial Intelligence: Methodology, Systems, and Applications.
- Balint, M., Dascalu, M., & Trausan-Matu, S. (2016b). *The rhetorical nature of rhythm*. Paper presented at the RoEduNet Conference: Networking in Education and Research, 2016 15th.
- Drory, R. (2000). *Models and contacts: Arabic literature and its impact on medieval Jewish culture* (Vol. 25): Brill.
- Fernandez, R. (2004). *A computational model for the automatic recognition of effect in speech*. Massachusetts Institute of Technology.
- Jamal, N., Mohd, M., & Noah, S. A. (2012). Poetry classification using support vector machines. *Journal of Computer Science*, 8(9), 1441.
- Jayyusi, S. K. (1977). *Trends and movements in modern Arabic poetry* (Vol. 6): Brill.
- Kaplan, D. M., & Blei, D. M. (2007). *A computational approach to style in American poetry*. Paper presented at the Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on.
- Lefer, M.-A. (2011). Contrastive word-formation today: Retrospect and prospect. *Poznań Studies in Contemporary Linguistics PSiCL*, 47, 645.
- Martin, J. H., & Jurafsky, D. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*: Pearson/Prentice Hall.
- Nehar, A., Ziadi, D., Cherroun, H., & Guellouma, Y. (2012). *An efficient stemming from Arabic text classification*. Paper presented at the Innovations in Information Technology (IIT), 2012 International Conference on.
- Niculescu, I.-D., & Trausan-Matu, S. (2016). *Rhythm analysis of texts using Natural Language Processing*. Paper presented at the RoCHI.

- Rakshit, G., Ghosh, A., Bhattacharyya, P., & Haffari, G. (2015). *Automated Analysis of Bangla Poetry for Classification and Poet Identification*. Paper presented at the Proceedings of the 12th International Conference on Natural Language Processing.
- Saleh, A.-Z. A. K., & Elshafei, M. (2012). Arabic poetry meter identification system and method: Google Patents.
- Tizhoosh, H. R., & Dara, R. A. (2006). On poem recognition. *Pattern analysis and applications*, 9(4), 325-338.
- Tizhoosh, H. R., Sahba, F., & Dara, R. (2008). Poetic features for poem recognition: A comparative study. *Journal of Pattern Recognition Research*, 3(1), 24-39.
- Trausan-Matu, S. (2012). *Repetition as artifact generation in polyphonic CSCL chats*. Paper presented at the Emerging Intelligent Data and Web Technologies (EIDWT), 2012 Third International Conference on.
- van Gelder, G. J. (2012). *Farrin, Abundance from the Desert: Classical Arabic Poetry*. (Middle East Literature in Translation.) Syracuse, NY: Syracuse University Press, 2011. Pp. xix, 364. \$24.95. ISBN: 9780815632221.