International Journal of User-System Interaction 17 (3) 2024, 65-84

Translating the Pulse of Poe: A Stylistic Comparison of Rhythm in English and Romanian Versions

Florin-Cristian Motrun[1], Ștefan Trăușan-Matu[1,2,3]

[1] National University of Science and Technology Politehnica Bucharest
313 Splaiul Independenței, Bucharest, Romania

[2] Research Institute for Artificial Intelligence
13 Calea 13 Septembrie, Bucharest, Romania

[3] Academy of Romanian Scientists
54 Splaiul Independenței, Bucharest, Romania
E-mail: florin.motrun@stud.acs.upb.ro, stefan.trausan@upb.ro

Abstract. The paper analyzes the rhythmic profile of Edgar Allan Poe's poems together with Emil Gulian's Romanian translations. It is presented as a pipeline that integrates three Romanian mathematical models developed by Solomon Marcus, Mihai Dinu, and Vasile Vasile, with contemporary natural language processing methods to automate syllable extraction, stress assignment, and rhythmic pattern detection. Repetition phenomena and additional rhythmic devices are also quantified. Applying the system to aligned poem pairs reveals how Poe's original rhythmic structures are preserved, amplified, or reshaped in translation, providing complementary quantitative metrics and qualitative insights into cross-lingual poetic rhythm.

Keywords: rhythm analysis, rhythmic devices, stress, syllable, natural language processing, translation

DOI: 10.37789/ijusi.2024.17.3.1

1. Introduction

In recent years, natural language processing (NLP) became a major technology for human-computer and even human-human interaction, especially due to the astonishing achievements of deep learning in developing large language models. However, in the case of the generation and translation of poems, as compared to human performance, some problems appear (De la Rosa et al., 2023; Resend and Hadley, 2024) due to their specific aesthetic

and emotional features. If musicality is intrinsically related to natural language, in general (Trausan-Matu, 2020), in the case of poetry is fundamental, rhythm being a major ingredient.

Rhythm may arise from repeated words (Niculescu and Trausan-Matu, 2016) or from the regular alternation of stressed and unstressed syllables. It enhances expressiveness and clarity, reinforces meaning, establishes pacing, and lends a musical quality that makes the text more engaging, memorable, and easier to comprehend.

In both Romanian and English, rhythm depends on accurate syllabification and stress (accent) placement. While humans easily segment syllables by jaw movement, algorithmic syllabification must handle numerous irregularities beyond simple rule sets. English words often violate straightforward splitting rules, while Romanian's vowel combinations and compound structures also produce exceptions. Robust syllabification algorithms must therefore integrate multiple phonotactic constraints and exception lists to achieve the precision necessary for reliable rhythm analysis.

The accent is the phonetic prominence given to one syllable within a word. In English, stress patterns are irregular - while many two-syllable nouns favor penultimate stress, countless exceptions exist - so automatic stress assignment can vary significantly from one word to another. In Romanian, most words carry stress on one of the last two syllables, yet fixed rules do not cover all cases, making rule-based accent determination unreliable in most cases.

To address both challenges, we leverage two large-scale lexical resources. For English, the CMU Pronouncing Dictionary (Carnegie Mellon University, 2014) provides over 134.000 entries with syllable boundaries and stress annotations. For Romanian, RoLEX (Lőrincz et al., 2023) offers 330.000 lemmas complete with part-of-speech tags, syllabification, lexical stress, and phonetic transcription - its neural-network predictions trained on smaller RoSyllabiDict and MaRePhoR corpora.

For rhythm analysis, there were used mathematical formalisms by Solomon Marcus (Marcus, 1970) and Mihai Dinu (Dinu, 1986), alongside Vasile Vasile's study of euphonic vowel-cluster repetition in Eminescu's verse (Vasile, 2005). In addition, pattern-matching algorithms to detect rhythmic devices were integrated.

The aim of this research is to build a precise, end-to-end pipeline that combines large-scale lexical resources, mathematical rhythm models, and automated detection of repetition figures to analyze the prosody of Edgar

Allan Poe's poems alongside Romanian translations. This will enable more rapid and effective insights into their rhythmic structure, automating what is currently a time-consuming manual process and yielding quantitative insights into how poetic form transfers across languages.

The paper reviews the rhythmic theory and the works of Edgar Allan Poe, surveys English and Romanian digital dictionaries, and presents three mathematical models of rhythm alongside common rhythmic devices. It proceeds to detail data processing, the implementation of each analysis method and device detection, and finally presents the results and conclusions.

2. State of the Art

2.1. Computational and Quantitative Approaches to Poetic Rhythm

Recent work in computational poetic analysis focuses on automating the extraction of rhythmic features such as stress patterns and metrical structure from poetry corpora. Greene, Bodrumlu, and Knight (2010) present a statistical method that automatically analyses rhythmic patterns in English verse by learning word-stress patterns from raw poetry and applying them to generation and translation tasks. Their system connects stress analysis directly to rhythmic structure, demonstrating the potential and challenges of unsupervised stress detection in poetic text.

Similarly, artificial intelligence, transformer-based language models have been evaluated for automatic metrical pattern prediction across languages, indicating that large pretrained models can capture rhythmic information and enable cross-lingual prosody analysis. This work suggests that multilingual NLP models may generalize rhythmic pattern recognition beyond a single language, which is especially relevant for cross-lingual rhythm comparison (De la Rosa et al., 2023).

2.2. The RoLEX Dictionary

RoLEX is currently the most comprehensive validated phonetic lexicon for Romanian, containing 330.866 entries. Each word is linked to its lemma, MSD tag, syllabification, lexical accent, and phonetic transcription. Initial

predictions for phonetic segments, syllable boundaries, and accent were generated by a sequence-to-sequence neural front end and then manually verified. Evaluation shows that, when trained on a carefully selected subset, word-level error rates drop to 3%, demonstrating that AI models can generalize well from relatively small data. Today, RoLEX underpins Romanian Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) systems and is published as Linked Open Data - queryable via SPARQL for easy integration with other semantic resources (Lőrincz et al., 2023).

2.3. The CMU Pronouncing Dictionary

CMUdict, developed by Carnegie Mellon's Speech Group in the 1990s, provides over 134.000 North American pronunciations in ARPABET - a 39-symbol phoneme set encoding consonants, vowels, and stress levels. Since version 0.7b (2014) it has been open source and serves as the reference corpus for training grapheme-to-phoneme (G2P) systems (Carnegie Mellon University, 2014). CMUdict remains a central resource in AI pipelines for ASR, TTS, and rhyme generation.

2.4. Rhythm and Musicality in Poe's Poetry

Poe treats his poems as sound scores - "sound precedes sense" (Poe, 1846) - each text being built on a precise rhythmic engine, enhanced by euphony, refrains, and onomatopoeia. His poetry occupies a distinctive position in the nineteenth-century due to its systematic emphasis on sound, rhythm, and musical organisation as primary carriers of aesthetic and emotional meaning. Poe explicitly articulated this priority in "The Philosophy of Composition" (Poe, 1846), where he argued that poetic effect arises from deliberate orchestration of meter, sound repetition, and refrain rather than from narrative alone. Subsequent literary scholarship has confirmed that Poe's verse is particularly amenable to formal rhythm analysis, as its prosodic structures are highly patterned, foregrounded, and often self-conscious (Miller, 2020; Attrige, 1995). This study focuses on four poems - The Raven, Annabel Lee, Ulalume, and The Bells - selected because they are canonical works in Poe's oeuvre and because each foregrounds rhythm and musicality in a distinct and analytically meaningful way. Together, they form a compact yet representative corpus illustrating the diversity of Poe's rhythmic techniques, making them suitable for both qualitative and computational analysis.

The Raven (1845) is widely regarded as Poe's most formally controlled poem and has been repeatedly analyzed for its *highly regular trochaic octameter*, internal rhyme, and refrain-based structure. Scholars have emphasized that the poem's obsessive rhythm - reinforced by alliteration, caesurae, and the repeated refrain "Nevermore" - plays a crucial role in shaping the reader's emotional experience, mirroring the speaker's psychological fixation (Creanga, 2022; Ferber, 2019). Its extreme regularity and dense sound patterning make *The Raven* a benchmark text for rhythm-focused stylistic analysis and for evaluating how such structures behave in translation.

Annabel Lee (1849) exemplifies a more lyrical and melodic form of rhythm construction. Critical analyses note that its anapestic tendencies, end-rhymes and repeated syntactic frames ("in this kingdom by the sea") create a song-like cadence reminiscent of ballads and oral poetry traditions. The poem's rhythm functions not only as a structural device but also as a mnemonic and affective mechanism, reiforcing themes of idealized love and loss through repetition and sonic echo (Whissell, 2011). This makes *Annabel Lee* particularly relevant for examining how rhythm supports emotional continuity across languages.

Ulalume (1847) represents Poe's most sound-driven poem and is often cited as an extreme case of phonetic orchestration and incantatory rhythm. Originally composed for elocution, the poem foregrounds alliteration, assonance and cyclic rhythmic patterns to such an extent that sound frequently dominates semantic clarity. Literary critics have argued that *Ulalume* exemplifies Poe's conception of poetry as an auditory experience, where rhythmic flow guides interpretation and emotional immersion (Miller, 2020). Its dense and unusual sound patterns provide a challenging test case for computational rhythm analysis and for studying rhythmic transformation in translation.

The Bells (1849) occupies a unique place in Poe's poetic corpus due to its overt imitation of acoustic rhythm and sound movement. The poem famously employs rapid shifts in meter, extensive repetition, and onomatopoeia to simulate the ringing of different types of bells, progressing from light, rapid rhythms to heavy, tolling cadences. Scholars have described *The Bells* as an experiment in rhythmic excess, where meter, repetition, and phonetic density function almost independently of narrative content (Attridge, 1995; Miller, 2020). Its extreme variability and rhythmic escalation make it especially

valuable for analyzing dynamic rhythm patterns and their computational detectability.

Taken together, these poems form a deliberately curated corpus that captures Poe's rhythmic range: from strict metrical regularity (*The Raven*) to lyrical repetition (*Annabel Lee*), to incantatory sound dominant verse (*Ulalume*), and finally to rhythm acoustic simulation (*The Bells*). Their selection is therefore not arbitrary but grounded in literary scholarship that recognized rhythm and musicality as defining features of Poe's poetic practice. This makes them particularly appropriate for the present study's goal of combining mathematical rhythm models and computational analysis to explore how poetic rhythm is preserved, reshaped or reinterpreted across languages.

- **Vowel distribution and euphony**

In the first stanza of "Ulalume" open (a), half-open (o, ă, e), and closed (u, î) vowels alternate to create a slow, hypnotic "breathing" effect (see Table 1).

Table 1. Vowel counts in "Ulalume" in the Romanian version

Line	Open	Half-open	Closed
Cerurile erau sure şi aspre	2	5	5
Frunzele erau crispate şi seci	2	5	5
Frunzele erau istovite şi reci	1	6	6
Era noapte în pierdutul Octombrie	2	7	5

The surplus of half-open vowels gives an autumnal haze, while repeated "s-" sounds ("sure", "aspre") reinforce the atmosphere - a clear example of how Poe's attention to sound shapes emotional effect (Whissell, 2011).

- **Repetition and counterpoint**

Poe's monosyllabic refrains ("Nevermore", "Lenore") anchor an intense emotional effect, while the use of parallel rhymes (e.g., "as-spoken" / "soul-token" in "The Raven", stanza 15) generates a layered internal echo. This pattern of recurring sounds contributes to what Creangă (2022) describes as

a verbal orchestration, in which rhyme, alliteration, and repeated phonetic textures work together to produce a structured - almost musical - play of repeated elements, a kind of sonic counterpoint.

- **Metric variety and hemiola**

In "The Raven", the catalectic trochaic octameter ("Once u|pon'| a mid'|night drear'|y while'| I pon'|dered wea'|ry'') drops the final foot to produce a rhythmic "fall". In "The Bells*"*, Poe mixes anapests and trochees, imitating a bell's swing and its clamor - a rhythm shaped by overlapping sound patterns, creating a hemiola-like effect (Miller, 2020). The hypnotic meter of "Annabel Lee" combines catalectic iambs and anapests, reinforced by internal rhymes ("sea/Lee/me"), producing a rocking, lullaby-like rhythm.

Through vowel patterns, alliteration, and metrical design, Poe elevates rhythm to a semantic force: sound itself - persistent, alliterative, anaphoric - drives the reader's emotional "unity of effect". Modern studies confirm that Poe's tension of pattern and deviation anticipates free verse (Attridge, 1995) and that his repetitions can be generated algorithmically (Fabb and Halle, 2008).

2.5. The method of Solomon Marcus

Solomon Marcus's approach (Marcus, 1970) models a text's rhythm via its "rhythmic structure" - the sequence of counts of unstressed syllables between stressed ones - then derives key metrics:

- **Rhythmic structure:** derived from the positions of stresses: for each pair of successive accents, count the number of unstressed syllables between them, including the second stressed syllable.
- **Rhythmic length:** the number of stresses in a phrase (length of the structure).
- **Rhythmic index:** a mathematical ratio measuring the density of stresses in a line, computed as the smallest k for which the total number of unstressed syllables does not exceed k times the number of stresses.
- **Upper/lower margins:** the maximum/minimum counts in the structure.
- **Rhythmic diameter:** the difference between upper and lower

margins.

Together, these quantify how uniformly or variably a text employs its stressed-to-unstressed patterns.

2.6. The method of Mihai Dinu

This method (Dinu, 1986) consists in segmenting each line into mono-accent units (u), each containing exactly one stressed syllable and up to six unstressed ones.

Dinu computes for each unit:

- **l(u):** number of syllables.
- **a(u):** position of the stressed syllable.
- **$d(u_i, u_{i+1})$:** gap between successive stresses.

Additional key metrics:

- **Line measure:** the sum of the lengths of the rhythmic units in a line.
- **Meter type** ($n \in \{2, 3, 4\}$): chosen if all gaps d are divisible by n, yielding binary, ternary, or quaternary pulse.

In English, the prevalence of neutral vowels, secondary stresses, and syllable contractions breaks the one-stress-per-unit assumption, making the method unstable without normalization.

2.7. The method of Vasile Vasile

Vasile Vasile's approach models poetic musicality via periodic vowel triplets (Vasile, 2005). Sliding a window of three successive vowels, he classifies each triplet by majority:

- **Open (a, ă, e, o):** if at least two of the three vowels belong to this group, yielding a bright, resonant timbre.
- **Closed (i, î/â, u):** if at least two vowels are from this set, producing a dark, tense timbre.

A verse acquires rhythm when these triplets recur at regular intervals and especially when they alternate between open and closed registers. This alternation intensifies musicality and semantic contrast. In English, however, vowel reduction and unstable diphthongs weaken these timbral contrasts,

limiting the method's applicability.

2.8. Rhythmic Devices

Rhythmic devices range from metrical feet at the micro level to word-repetition figures at the lexical-syntactic level, each imparting a perceivable pulse to the verse. While feet establish the basic meter, targeted repetitions of words or phrases in specific positions generate additional stresses and semantic echoes, deeply enhancing a poem's emotional impact (Malyshkina et al., 2018; Lagutina et al., 2020).
These repetitions, appearing at key moments, include:

- **Anaphora:** repeating a word or phrase at the beginning of successive units to link and emphasize themes.
- **Epiphora:** repeating at the end of successive units to create an echo and reinforce conclusions.
- **Symploce:** combining anaphora and epiphora in adjacent lines with matching openings and closings for a framed effect.
- **Anadiplosis:** starting a line with the word or phrase that ended the previous line, boosting cohesion.
- **Epanalepsis:** repeating a term with slight variation in close proximity, adding an internal pulse.
- **Reduplication:** immediate doubling of a word or expression to intensify emotion.
- **Epistrophe:** repeating within the same verse to establish periodicity and memorability.
- **Polysyndeton:** using extra conjunctions to slow the rhythm and add solemnity.
- **Aposiopesis:** abruptly breaking off a line to introduce syncopation and suspense.

3. Implementation

3.1. The language and libraries used

The implementation uses Python for its adequacy for NLP and rich ecosystem. Key libraries include:

- **pathlib, csv, collections:** file I/O and data structures (Counter, defaultdict).
- **re:** flexible, high-performance text extraction via regular expressions.
- **nltk:** tokenization and access to the CMU Pronouncing Dictionary for English phonetics.
- **spaCy & stanza:** language models (en_core_web_sm, ro_core_news_sm, UD) for POS tagging and parsing.
- **pyphen:** hyphenation using Hunspell dictionaries (en_US, ro) for syllabification.
- **RoLEX:** Romanian lexicon providing syllable splits and stress positions.

By combining regex extraction, RoLEX for Romanian stress, CMUdict for English phonetics, hyphenation for out-of-vocabulary syllables, and advanced NLP via spaCy/stanza, it is achieved a full, bilingual rhythmic analysis pipeline.

3.2. Functionalities

The workflow of the application, including the processing modules, is presented in Figure 1. The details of each processing step and modules are discussed below.

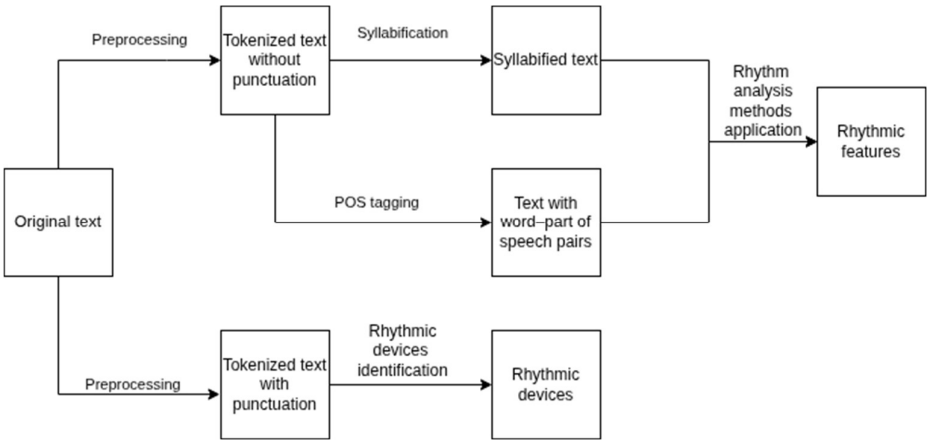

Figure 1. The workflow of the application

1) Data Preparation Workflow

Before applying the rhythm analysis and rhythmic methods, the following steps are performed to ensure clean and structured data:

a) Text preprocessing: The text is read from file, skipping metadata: title, author, year. Lines are trimmed, and empty rows or newline characters are removed.

b) Tokenization:

- **Romanian:** Tokenization is performed using a regular expression (Motrun, 2025). Diacritics and hyphenated or apostrophized words are preserved. Punctuation and digits are removed, and text is lowercased.
- **English:** It is used *wordpunct_tokenize()* from the **NLTK** library, followed by lowercasing and removal of punctuation and digits.

c) Syllabification:

- **Romanian:**
  - *Resources:* Preprocessed file and **RoLEX** dictionary.
  - *Phonetic categories:* vowels, consonants, diphthongs, triphthongs, inseparable clusters (e.g., che, ge) and consonant group patterns (e.g., lpt, nct).
  - *Rules:*
    - Handle hyphenated words (e.g., short prefixes/ suffixes are joined).
    - Use RoLEX if the word exists in the dictionary.
    - If the word is not found in RoLEX, apply *pyphen* or *fallback* to manual rules: identify diphthongs and triphthongs, apply consonant splitting logic (1, 2, 3+ consonants), and use buffer-based parsing from left to right.
    - *Output:* Each syllabified word is saved in CSV format with hyphens; hyphen-linked fusions are marked in parentheses.
- **English:**
  - *Resources:* Preprocessed file and **CMU Pronouncing Dictionary** (via NLTK).

- *Phonetic categories:* vowels, valid/invalid consonant clusters, indivisible suffixes, initial mute clusters (e.g., kn), indivisible roots (e.g., qu).
- *Two methods:*
  - **syllabify_cmu()**: Uses CMUDict if available; otherwise returns whole word or splits syllables proportionally.
  - **syllabify_manual()**: Applies fallback rules like blocking mute clusters, separating indivisible affixes, applying the Maximum Onset Principle, and splitting double consonants.
- *Fallback logic:*
  - Apply syllabify_cmu();
  - If not found in CMUDict, use pyphen;
  - Finally apply syllabify_manual().
- *Output:* All syllabified words saved in CSV format with hyphen separators.

d) Part-of-speech tagging: Words are tagged with their part of speech and stored as word/tag pairs. For Romanian, tagging is performed using **Stanza** with the Romanian model. For English, tagging uses **spaCy** with the *en_core_web_sm* model.

This structured workflow provides consistent and language-adapted input data for the rhythm analysis methods that follow.

2) Accent Assignment

Stress (accent) assignment is essential for rhythm identification. Figure 2 presents the implied processing.

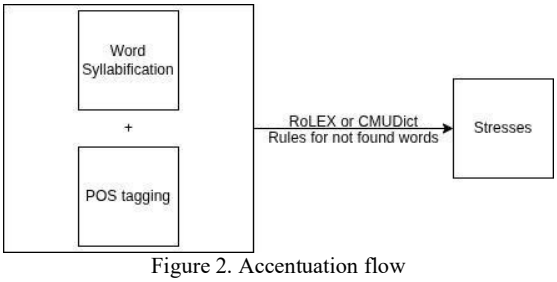

Figure 2. Accentuation flow

- **Romanian:** POS tags from Stanza are mapped to RoLEX categories. Each syllabified word is paired with its POS tag. If found in RoLEX, the stress pattern is extracted; otherwise, in words of two or more syllables, the stress is on the penultimate syllable. Monosyllabic words are stressed on their single syllable. All syllables and stresses are concatenated, and accent positions and syllable counts are computed.
- **English:** It is used the word, its POS tag, and syllable count to determine stress: helper words are unstressed, CMUDict words use dictionary stress, others follow fallback rules. Syllables and accents are concatenated and analyzed for stress positions and total count.

3) Implementation of the Solomon Marcus' method

For each verse are computed the number of syllables and accents, the accent positions, the rhythmic structure and length, the upper/lower margins, the rhythmic diameter and index. Additionally, frequency distributions are computed for the number of syllables, the rhythmic length, the lower and upper rhythmic edges, the rhythmic diameter and index.

4) Implementation of the Mihai Dinu's method

For each verse are computed the number of monoaccentual units (each consisting of a single stressed syllable and optional unstressed ones), the length of each unit (in syllables), the number of syllables in each rhythmic unit, the accent positions within each unit, the distances between stressed syllables in consecutive units, the rhythm type, based on divisibility by 2, 3, or 4 (applied when ≥ 50% of units satisfy the condition), `the` measure (total number of syllables per verse) and the structure type (acatalectic or catalectic). Additionally, frequency distributions are computed for the unit lengths and ranks, distances, rhythm types, measures and number of units of each verse.

5) Implementation of the Vasile Vasile's method

For Romanian, as a general logic, for each verse:

- All vowels are extracted in order of appearance.
- Successive triplets are built: v1-v2-v3, v4-v5-v6, etc.

- Each triplet is classified based on the number of vowels from the defined registers.

Special cases handling:

- If the total number of vowels per verse is a multiple of three plus one: ignore the 4th vowel.
- If the total number of vowels per verse is a multiple of three plus two: the last 5 vowels are grouped into two overlapping triplets to ensure complete coverage. The last two triplets are the triplet formed by the last three vowels and the triplet formed by the three vowels that precede the final two vowels.

English adaptation: Although the open/closed contrast is less explicit in English, vowels and diphthongs are grouped by timbral register based on tongue height, jaw openness, and vocal tension (see Table 2). This classification allows applying the same logic as in Romanian. Each word is first converted into its phonetic transcription using the CMU Pronouncing Dictionary, which outputs a sequence of phonemes in ARPABET format. From this transcription, only vowel and diphthong symbols are extracted and grouped into triplets.

Table 2. Timbral Register Classification for English Vowels

Register	ARPABET Symbols
Open / Semi-open	AA, AE, AH, AO, AW, AY, EH, EY, OW
Closed	IY, IH, UW, UH, ER, OY, YU

Classification rules for English:

- Low and mid-low vowels ⇒ Open.
- Mid-high and high vowels ⇒ Closed.
- Diphthongs are classified by their final component.

For each verse, are computed the list and type of all triplets, the number of open and closed groups, the open/closed ratio, and the alternation count between group types.

6) Rhythmic devices identification

For each rhythmic figure, a dedicated search function is implemented to

scan the poem line by line. The main detection logic is:

- **Anaphora:** identifies up to 3-word identical openings in at least two verses; returns all non-overlapping groups with verse indices.
- **Epiphora:** same as anaphora but applied at the end of the verses.
- **Symploce:** combines anaphora and epiphora results; retains only the verses where both are simultaneously present.
- **Anadiplosis:** for each pair of consecutive verses, checks if the end of the first matches the beginning of the second (up to 3 words); returns matching pairs with indices.
- **Epanalepsis:** searches for word/group repetition within a verse at a distance of 2–3 words; only the first match per verse is kept.
- **Reduplication:** detects immediate repetition of a word/group (up to 3 words) within the same verse; returns the first such sequence.
- **Epistrophe:** finds repeated word/group occurring at a distance of 4 or more words in the same verse; returns the matched phrase and verse index.
- **Polysyndeton:** counts conjunctions in a verse; if at least two are found and they constitute at least 20% of all words in the verse, the verse is flagged.
- **Aposiopesis:** detects punctuation-based interruptions (e.g., …) using regex; operates on the punctuation-preserved text.

3. Results

The study was conducted on four poems by Edgar Allan Poe - Annabel Lee, The Bells, The Raven, and Ulalume - along with their Romanian translations by Emil Gulian (1938), using a combination of computational rhythmic analysis and structural pattern detection.

3.1. Solomon Marcus Method

The Romanian translations show a clear tendency toward *metric regularity*: over 70% of verses have between 10 and 12 syllables, with frequency peaks at 10 or 11 syllables. In contrast, the English originals display a broader range, from 7 to 16 syllables, with 7, 8, and 9 being the most frequent lengths.

The *rhythmic index* reaches its maximum at 9–10 syllables in Romanian

and at 7–9 syllables in English, highlighting a concentration of well-balanced accent distribution at those lengths. The *rhythmic diameter*, reflecting the dispersion of accents, remains moderate in Romanian (generally between 30-41%), while in English it can reach significantly higher levels, up to 80.5% for shorter lines. The *upper and lower rhythmic margins* fall within a consistent range of 24-47% in both languages. Exact repetition of *rhythmic structure* per verse is rare, under 1% in both languages.

At the *syllable level*, accent placement alignment between original and translation occurs in 23-53% of the syllables in English, and slightly lower in Romanian, between 23-31%.

3.2. Mihai Dinu Method

The number of *rhythmic units* per verse (as a macro-unit of measure) rarely coincides between English and Romanian, with matching rates ranging from 0 to 12%.

The *verse length* in syllables shows slightly better alignment, with 4-17% of verses having identical counts. However, the matching of *rhythmic measures* - understood as the number of full metrical feet - shows higher consistency, occurring in 31-48% of verses.

The *structural type* (acatalectic vs. catalectic) matches in 44-51% of verses, while the overall *rhythmic type* (binary, ternary, quaternary) reaches its best correspondence in "The Raven" (up to 82%) and ranges between 40-56% for the other texts.

In terms of *accent rank*, Romanian translations tend to emphasize strong initial accents (rank 1) in 43-50% of verses, whereas the English originals more often use rank 2 or 3 (33-50%), reflecting a more medial accentuation.

The *distance between accents* also differs: Romanian favors compact sequences with 1-2 syllables between accents (with up to 43.3% at distance 1), while English typically features wider spacing of 2-3 syllables, with distance 2 being the most frequent (up to 55.4%).

3.3. Vasile Vasile Method

The *phonetic structure of vowel triplets* reveals consistent contrasts between the two languages. In the Romanian texts, between 65% and 73% of vowel triplets are open, while English originals tend to have an even higher rate, often between 76% and 81%.

Romanian displays a broader dispersion, with the *top 5 most frequent triplets* accounting for only 14% of all cases, indicating high diversity. In English, the most common triplets are slightly more concentrated, each covering around 3-6%.

The *alternation values* - indicating changes between open and closed vowels - range from 1.4 to 2.7 (mean) and 0.3 to 0.34 (relative) in Romanian, while English texts show lower alternation levels: between 0.47 and 0.9 (mean), and 0.15 to 0.25 (relative).

The *open/closed group type matching* is more consistent in English, ranging between 63% and 70%, compared to only 41% to 47% in Romanian. Similarly, matching of *closed groups* is slightly better in English (6-10%) than in Romanian (4-7%).

For *full triplet group alignment*, English once again shows higher consistency, with 56-63% of matched triplets, while Romanian versions stay within the 35-46% range.

The *ratio between open and closed* triplets is more stable across English poems (50-70%) than in Romanian, where it shows higher variability from text to text.

3.4. Rhythmic Devices

Across all texts, the most frequently detected device was *aposiopesis*, with high counts in both languages (e.g., 39 in English and 44 in Romanian for Ulalume). One the other side, *symploce* is relatively rare, with low counts across all poems.

Epiphora appears consistently (up to 26 occurrences in English, 24 in Romanian), while *epanalepsis* occurs more often in Romanian in certain poems (e.g., The Raven: 5 English vs. 9 Romanian). *Anaphora* is most frequent in Ulalume (14 in English, 6 in Romanian).

Reduplication is notably more frequent in English (e.g., 23 in The Bells) and nearly absent in Romanian. *Polysyndeton* shows slight variation (e.g., The Bells: 2 English, 4 Romanian). *Anadiplosis* is concentrated in The Bells (11 English, 10 Romanian), and absent elsewhere. *Epistrophe* occurs in all poems but with varied frequency (e.g., 14 in English and 19 in Romanian in The Raven).

4. Conclusions

Romanian translations of Poe's poems exhibit greater *metrical regularity* and *accent cohesion*, typically centered around 10-12 syllables, while the English originals display *freer rhythmic variation* with peaks at 7-9 syllables (Marcus Method).

According to Dinu's model, Romanian versions favor binary rhythm and initial stress, while English alternates between binary and ternary patterns with more *balanced accent placement*.

Vasile's approach highlights that both languages prefer open vowel triplets, yet English maintains *smoother tonal consistency*, while Romanian explores a *phonemic variation*.

The analysis of rhythmic devices shows that auditory or visually expressive figures (e.g., aposiopesis, epiphora, epanalepsis) are better preserved in translation, whereas lexically dependent repetitions (e.g., anaphora, symploce, reduplication) are often lost or adapted.

The comparative analysis of the four poems and their Romanian translations reveals systematic differences in rhythmic structure, prosodic balance, and phonetic texture between the two languages:

- English originals show greater consistency in metrical parameters - syllable counts, accent distances, and rhythmic indices are more stable.
- Romanian translations, constrained by fixed stress and lexical differences, display lower micro-level fidelity (accent position, syllabic alignment) but maintain overall *prosodic balance* and rhythmic structure types.
- In terms of rhythmic units (Dinu), macro-structural features such as measure and rhythm type are preserved in 40-50% of verses, while accent-level accuracy drops below 20%.
- Vocalic patterns (Vasile) demonstrate higher triplet and alternation consistency in English, whereas Romanian translations are less regular but preserve phonetic richness.

This study confirms that the three computational methods capture complementary layers of prosody: Solomon Marcus quantifies macro-accentual regularity; Mihai Dinu maps metrical and rhythmic organization; and Vasile Vasile reveals phonetic rhythm through vocalic triplets.

Together, they show that while translations do not reproduce every metric nuance, they preserve the essential rhythmic structures and prominent auditory effects. Therefore, the methods prove both linguistically sensitive and methodologically valuable for comparative poetic analysis across languages.

References

Attridge, D. (1995). *Poetic rhythm: An introduction*. Cambridge University Press. https://www.academia.edu/6906196/Poetic_rhythm_An_introduction, accessed on 19 November 2025

Carnegie Mellon University. (2014). The CMU Pronouncing Dictionary (Version 0.7b) [Data set]. *Carnegie Mellon University*. http://www.speech.cs.cmu.edu/cgi-bin/cmudict, accessed on 19 November 2025

Creangă, M.-T. (2022). The Translatability of Poetry: Phonaesthesia, Sound Iconicity, Orchestration, and Aesthetic Function. A Case Study of Poe's The Raven. *American British and Canadian Studies*, 38(1), 163–178. https://doi.org/10.2478/abcsj-2022-0009

De la Rosa, J., Pérez, Á., De Sisto, M., Hernández, L., Díaz, A., Ros, S., & González-Blanco, E. (2023). Transformers analyzing poetry: multilingual metrical pattern prediction with transfomer-based language models. *Neural Computing and Applications*, *35*(25), 18171-18176.

Dinu, M. (1986*). Ritm și rimă în poezia românească.* Cartea Românească.

Fabb, N. and Halle, M. (2008). *Meter in poetry: A new theory*, Cambridge University Press. https://doi.org/10.1017/S0952675710000291

Ferber, M. (2019). *Poetry and language: The linguistics of verse*. Cambridge University Press.

Greene, E., Bodrumlu, T., & Knight, K. (2010, October). Automatic analysis of rhythmic poetry with applications to generation and translation. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 524-533).Lagutina, N. S., Lagutina, K. V., Boychuk, E. I., Vorontsova, I. A., & Paramonov, I. V. (2020). Automated rhythmic device search in literary texts applied to comparing original and translated texts as exemplified by English to Russian translations. *Automatic Control and Computer Sciences*, 54(7), 697-711. https://doi.org/10.3103/S0146411620070147

Lőrincz, B., Irimia, E., Stan, A., & Barbu Mititelu, V. (2023). RoLEX: The development of an extended Romanian lexical dataset and its evaluation at predicting concurrent lexical information. *Natural Language Engineering*, 29(3), 720-745. https://doi.org/10.1017/S1351324922000419

Malyshkina, E. V., Goroshko, O. N., Shevchenko, E. M., & NA, L. (2018). Scientific understanding of the functions of repetition in English poetic speech. *Revista Espacios*,

39(21), 41-59. https://www.revistaespacios.com/a18v39n21/a18v39n21p41.pdf, accessed on 19 November 2025

Marcus, S. (1970). *Poetica matematică*. Editura Academiei.

Miller, P. (2020). Prosody, media, and the poetry of Edgar Allan Poe. *PMLA*, 135(2), 315-328. https://doi.org/10.1632/pmla.2020.135.2.315

Motrun, F.-C. (2025). *Analiza stilistică a textelor*, Proiect de Diplomă, Universitatea Națională de Știință și Tehnologie POLITEHNICA București

Niculescu, I. D. and Trausan-Matu, S. (2016). Rhythm analysis in chats using natural language processing. In *Proceedings of the International Conference on Human-Computer Interaction* (pp. 107-112). https://rochi.utcluj.ro/articole/5/RoCHI2017-Niculescu.pdf, accessed on 19 November 2025

Poe, E. A. (1846). The philosophy of composition. *Graham's Magazine,* 28, 163-167. https://anthologydev.lib.virginia.edu/exist/apps/LiC/work/Poe/poe-philosophy, accessed on 19 November 2025

Poe, E. A. (1938). *Poemele lui Edgar Poe — Traduse din limba engleză în forma lor, cu un studiu introductiv de Emil Gulian*, Fundația Pentru Literatură și Artă „Regele Carol II"

Resende, N. and Hadley, J. (2024) The Translator's Canvas: Using LLMs to Enhance Poetry Translation. In Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track), pages 178–189, Chicago, USA. Association for Machine Translation in the Americas.

Trausan-Matu, S. (2020) Connections between human music and natural language, in Milena Slavcheva (ed.) *Knowledge, Language & Models*, INCOMA Ltd. Shoumen, pp. 302-317, ISBN 978-954-452-062-5

Vasile, V. (2005). De la muzica firii și a sufletului la muzica sferelor. *Muzica în viața și creația lui Eminescu,* Editura: Petrion.

Whissell, C. (2011). To those who feel rather than to those who think: Sound and emotion in Poe's poetry. *International Journal of English and Literature*, 2(6), 149-156. http://www.academicjournals.org/article/article1379424964_Whissell.pdf, accessed on 19 November 2025

Bergmann, D., Stryker, C. (2024). *What is artificial general intelligence (AGI)?* https://www.ibm.com/think/topics/artificial-general-intelligence, accessed on 9 July 2025

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)* https://www.english-corpora.org/coca/, accessed on 9 July 2025

Fraser, K C., Dawkins, H., Kiritchenko, S.(2025). Detecting AI-Generated Text: Factors Influencing Detectability with Current Methods, Journal of Artificial Intelligence Research Vol. 82 (2025) 2233-2278, https://doi.org/10.1613/jair.1.16665

Gehrmann, S., Strobelt, H., Rush, A. (2019) *GLTR: Statistical Detection and Visualization of Generated Text* https://aclanthology.org/P19-3019/, accessed on 9 July 2025

IBM (2023). *Large language models*. IBM. https://www.ibm.com/think/topics/large-language-models, accessed on 9 July 2025

.