

Received March 15, 2021, accepted March 17, 2021, date of publication March 30, 2021, date of current version April 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069635

# Automated Metric Analysis of Spanish Poetry: Two Complementary Approaches

GUILLERMO MARCO<sup>1</sup>, JAVIER DE LA ROSA<sup>2</sup>, JULIO GONZALO<sup>1</sup>,  
SALVADOR ROS<sup>2</sup>, (Senior Member, IEEE), AND ELENA GONZÁLEZ-BLANCO<sup>3</sup>

<sup>1</sup>Research Group in NLP & IR, National Distance Education University, 28040 Madrid, Spain

<sup>2</sup>Laboratory of Innovation on Digital humanities (LINHD), National Distance Education University, 28040 Madrid, Spain

<sup>3</sup>IE School of Human Sciences and Technology, IE University, 28006 Madrid, Spain

Corresponding author: Guillermo Marco (gmarco@lsi.uned.es)

This work was supported in part by the Starting Grant Research Project Poetry Standardization and Linked Open Data (POSTDATA) through the European Research Council (ERC) under the Research and Innovation Program Horizon2020 of the European Union under Grant ERC-2015-STG-679528, and in part by the Spanish Government, Project Misinformation and Miscommunication in social media: bias (MISMIS-BIAS) under Grant PGC2018-096212-B-C32.

**ABSTRACT** The automatic metric analysis (commonly referred to as *scansion*) of Spanish poetry is not a trivial problem since it combines the nuances of the language, the different poetic traditions related to melodic patterns, and the personal stylistic preferences and intentions of the author. In this paper, we explore two alternative algorithmic approaches tailored to different applications scenarios. The first approach, Rantanplan, is a rule-based method that consists of four Natural Language Processing modules that work together to perform scansion and other related analysis: Part of Speech tagging, syllabification, stress assignment, and metrical adjustment. The second approach, Jumper, explores the possibility of performing scansion without syllabification, with a twofold purpose: to minimize the errors propagated in different parts of the linguistic processing pipeline (including the syllabification step), and to improve the efficiency of the process. Both systems outperform the state of the art and provide either a more informative solution (suitable, for instance, for teaching purposes) or a more efficient processing (when a correct scansion is all the linguistic knowledge required, as in scholar philological studies). The combined use of both systems turns out to provide a practical tool to clean-up manual annotation errors in corpora.

**INDEX TERMS** Natural language processing, digital humanities, automated metric analysis, scansion, corpus annotation, poetry, Spanish.

## I. INTRODUCTION

In recent years, several systems for automating parts of the literary analysis have emerged, enabling corpus linguistic approaches on poetry corpora that would otherwise need unmanageable amounts of expensive manual annotation. In this work, we focus on the automated metric analysis of Spanish poetry.

The literary analysis of a poem in Spanish involves the scansion of its verses. A verse is a sequence of stressed and unstressed syllables delimited by metrical pauses [1]. The distribution of the stresses on this sequence determines the metrical pattern. Therefore, the scansion of a verse consists in the extraction of these metrical patterns of stressed and

unstressed syllables. Example 1 shows a verse of eleven syllables whose metrical pattern is 2.6.10.<sup>1</sup>

(1) *Amigos, el amor me perjudica*  
*A-mi-gos-el-a-mor-me-per-ju-di-ca*  
11 syllables – metrical pattern: 2.6.10  
(Julio Martínez Mesanza)

Automatic scansion of Spanish poetry is not a trivial task, as it combines the rules of the language, the explicit and implicit rules and traditions related to melodic patterns, and the personal stylistic preferences and intentions of the author. In this paper, we explore two alternative algorithmic solutions tailored to different applications scenarios. The first

<sup>1</sup>When representing metrical patterns, it is common in the Spanish tradition to only give the positions of the stressed syllables, as opposed to the entire metrical pattern, e.g., 2.6.10 would be equivalent to - + - - - + - - - + -, where the sign + means stressed.

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott<sup>1</sup>.

approach, Rantanplan (evolved from [2]), is a rule-based method that consists of four Natural Language Processing (NLP) modules that work together to perform scansion and other related analysis: PoS tagger, syllabification, stress assignment, and metrical adjustment. The second approach, Jumper (evolved from [3]), explores the possibility of performing accurate scansion without syllabification, with a double purpose: minimize the errors propagated in different parts of the NLP pipeline (including the syllabification step), and improve the efficiency of the process.

Rantanplan was designed to fit a wider scope of applications, as it provides much richer information: syllabification, phonological groups, rhyme, stanza types, position of rhetorical devices, and Part-of-Speech (PoS) tags, among others. Explicit syllabification can be particularly useful in teaching environments, helping students visualize and understand how metric analysis is performed. It also allows researchers to produce new results in processes such as close and distant reading and critical edition. That is possible since Rantanplan output is machine-readable, interoperable, and ready to be ingested into a Linked Open Data triple store compliant with the POSTDATA Project network of ontologies. [4].

Jumper, on the other hand, was designed to perform the scansion task with minimal NLP preprocessing while optimizing for accuracy and efficiency. Its output is well-suited for researchers, poets or translators, and it is related to metre: metrical pattern, position of rhetorical devices, verse type, and self-assessed confidence in the classification of the pattern.

Overall, both systems substantially improve previous state-of-the-art approaches (including rule-based and neural network-based methods), and provide optimal alternatives depending on the application scenario. In addition, our experimental results show that, when applied together, the comparative analysis of their outputs provide reliable hints to locate errors in the manual annotation process.

This paper is organized as follows. In Section II we discuss prior state of the art. In Section III we define the problem of scansion, the rhetorical devices that introduce ambiguity in metrical patterns extraction and how philologists usually solve the problem. In Sections IV and V, Rantanplan and Jumper algorithms are introduced. Subsequently, in Section VI we describe our experimental setting, discuss empirical results, and perform a detailed comparative error analysis. Finally, conclusions are drawn in Section VIII.

## II. RELATED WORK

There are approaches for the automated scansion of poetry that date back to the late 20th century, [5]–[7]. Unfortunately, not all languages and poetic traditions have received equal attention in developing automatic tools for poetry [8], with English being overwhelmingly overrepresented and with modest attempts in other languages such as German, France, and Spanish. To the best of our knowledge, Logan's work could be considered the first realization of these tools [6] for English. In his work, he developed a set of programs

to analyze sound and meter in poetry following the generative phonological theory of Chomsky. This first attempt was followed by *Scansion Machine*, a tool designed to analyze iambic pentameters [5] based on five pedagogical scansion steps [9]. A few years later, *Poetry Processor* was developed [7], a first commercial tool for scansion and designed to interactively assist users in composing metrical verse. At the same time, the first tool for the French language was documented: *Metromete* [10], a tool to explore different aspects under which rhythmic phenomena appear. *Metromete* was explicitly designed for the analysis of the classical alexandrine, a rigid and very restricted kind of verse. Scansion machine inspired *Scandroid*, a tool designed to analyze English verse in iambic and anapestic meters [11]. This system works at the verse level giving information about the number of syllables and their stress. For this task, a dictionary and the syllable division procedure built on the principles described by Newman [7] was used. However, it has a problem related to its approach to lexical stress, because there are lexical units that are composed by more than one word (e.g., phrasal verbs). Some other examples of these tools for English are *AnalysePoems* [12], *Calliope* [13], and *Prosodic* [14]. Only after the first decade of the twenty-first century did the tools devoted to scansion in different languages expand: *Gunstick* [15] for Czech, *Metricizer* [16], [17] for German, and De Sisto's work [18] for Dutch. Even a tool for Ancient Greek hexameter [19] was developed. A detailed summary of scansion tools in different languages can be found in [18]. For the Spanish traditions, although manuals on metrical analysis of Spanish poetry exist at least since the 18th century [20], the foundational work for modern analysis would take another century to appear (see, e.g., [21]–[23]). Despite such a long and rich tradition, it is not until 2000 that the first computational tool to assist scholars in the analysis of Spanish poetry was introduced by Gervás as part of a system for the automatic generation of metrical poetry [24]. In his work, Gervás used Prolog to model the division of a word into its constituent syllables, adding additional predicates to handle synaloepha and synaeresis. A more modern approach was introduced in 2017 by Navarro-Colorado [25]. He built a rule-based system leveraging the morphological analyzer in Freeling [25], [26] and focused on resolving metrical ambiguities. In his method, after splitting words into syllables and assigning stress according to their POS, the possible synaloephas and diaeresis are marked and applied, ignoring synaereses. This happens according to a knowledge base with probabilities for the different metrical patterns. The system reported accuracy of 94.44%. Besides, the authors reported an inter-annotator agreement rate of 96%. However, these results are restricted to hendecasyllabic verses. Shortly after that, Agirrezabal used neural networks to predict the metrical pattern of lines of verses [27]. The model proposed was a character-based bidirectional long short term (BiLSTM) neural network with conditional random fields. The per-line accuracy reported was 90.84%.

TABLE 1. Classification of eleven syllables verses [30, p. 118].

Endecasílabos		
— Heroicos	2.6.10	puro
	2.4.6.8.10	pleno
	2.4.6.10	corto
	2.6.8.10	largo
	2.4.10	difuso
— Melódicos	3.6.10	puro
	1.3.6.8.10	pleno
	3.6.8.10	largo
	1.3.6.10	corto
— Sáficos	4.8.10	puro
	1.4.8.10	puro pleno
	1.4.6.8.10	pleno
	4.6.10	corto
	1.4.6.10	corto pleno
	4.6.8.10	largo
	2.4.8.10	largo pleno
	4.10	difuso
	1.4.10	difuso pleno
	4.7.10	puro
— Dactílicos	1.4.7.10	pleno
	2.4.7.10	corto
	1.6.10	puro
— Enfáticos	1.6.8.10	pleno
	6.10	puro
— Vacíos	6.8.10	pleno
	1.4.10	pleno
	2.4.10	heroico

III. PROBLEM DEFINITION

Spanish orthography establishes rigid rules to assign stress and classifies the words according to the last stressed syllable position. There is generally only one stressed syllable per word,<sup>2</sup> with few exceptions [28]. Depending on the position of the stressed syllable, there are three categories of words:

- *oxytone* words, when the stressed syllable is the last syllable of the word: ‘tam-**bor**’.
- *paroxytone* words, when the stressed syllable is the one before the last syllable of the word: ‘**plan**-ta’.
- *proparoxytone* words, when the stressed syllable lies two syllables from the end of the word: ‘**plá**-ta-no’.

Similarly, according to its last stressed syllable, a verse is called oxytone, paroxytone or proparoxytone. From each of them, there are constant metrical phenomena, called *verse compensation*, that affect the number of syllables [29]:

- When a verse is oxytone, an extra syllable is counted.
- When a verse is paroxytone, the actual existing syllables are counted.
- When a verse is proparoxytone, one syllable less is counted.

Moreover, the Spanish poetry tradition has classified verses according to their metrical pattern. The number of possible combinations of stressed syllables is finite. For example, an hendecasyllable may have five accents or less, whose combinations give rise to the different types shown in Table 1.

<sup>2</sup>In this work we use hyphens as the syllabic separator for representation purposes, marking in bold the stressed syllable (e.g., ‘a-mo-**ro**-so’).

Following this criterion, each verse can be classified according to the proximity to each of the types: the characteristic stresses (also called *rhythmic stresses*) are taken into account to classify the type of verse; non-coincident stresses are treated as extra-rhythmic.

The extraction of the metric pattern is an ambiguous process. Verses could present more than one possible scansion because they are affected by some rhetorical devices that might alter the counting of stresses and even syllables present in verse, thus differentiating between metrical length and syllabic length.

The two most common of these figures in Spanish are synaloepha and synaereses. While both imply the union of separate phonological groups, the former acts between the last syllable of a word and the first of the next; for example in ‘la amaba’, ‘la’ and ‘a’ would be joined together. For the latter, the union occurs between the adjacent vowels within a word, ‘son-re-ír’ can be then split as ‘son-reír’ after synaereses. After applying these alterations, the number of syllables effectively shrinks for metrical purposes. On the other hand, dieresis is the metric phenomenon in which two vowels within the same syllable forming diphthongs are separated into different syllables, increasing the syllable count.

Also, verses above eleven syllables, called compound verses, are divided into semiverses (also called *hemistichs*). The scansion of the semiverses can be done separately. This separation introduces two metrical phenomena exclusive to this type of verse, frequent in mixed-metre poetry: firstly, according to its last stressed syllable, the syllable counting rules for oxytone, paroxytone and proparoxytone verses are applied in semiverses. Secondly, dyalepha is always applied at the boundary between semiverses. Example 2 shows a compound verse where these metrical phenomena occur.

(2) *Oh, qué frescor, qué música / de chopos de estación*  
 $14 = 8 - 1 / 6 + 1$   
 (Juan Ramón Jiménez)

Example 3 shows all possible scansions for the verse *Todas las tardes se muere un niño* by Federico García Lorca. Subexample 3 a only considers the synaloepha in the phonological group *reun*, resulting in a verse of 10 syllables whose metrical pattern is 1.4.7.9. However, this phonological group may present a dyalepha (Subexample 3 c) or appear in combination with a synaereses in the diphthong *ue* of *muere* (Subexamples 3 b and 3 d). The most sensitive cases are Subexamples 3 b and 3 c. Both are 11 syllables long but their metric patterns is different. So, which one is the correct scansion?

(3) *Todas las tardes se muere un niño*  
 (a) **To**-das-las-tar-des-se-mue-reun-ni-ño  
 metrical pattern: 1.4.7.9 – 10 syllables –  
 rhetorical devices: synaloepha

- (b) *To-das-las-tar-des-se-mi-e-re-un-ni-ño*  
metrical pattern: 1.4.8.10 – 11 syllables –  
rhetorical devices: synaloepha and dieresis
- (c) *To-das-las-tar-des-se-mue-re-un-ni-ño*  
metrical pattern: 1.4.7.10 – 11 syllables –  
rhetorical devices: dyalepha
- (d) *To-das-las-tar-des-se-mi-e-re-un-ni-ño*  
metrical pattern: 1.4.8.11 – 12 syllables –  
rhetorical devices: dyalepha and dieresis  
(Federico García Lorca)

Hispanic Philology scholars resolve metrical ambiguities relying on three criteria:

- 1) *Verse context*. If most of the poem lines are 11 syllables long, likely, the ambiguous line is also 11 syllables long.
- 2) *Traditional metrical patterns*. Poets aspire to write verses whose metrical pattern coincides with the most *beautiful* ones. Deviation from these patterns is considered an imperfection. In the case of ambiguity, it is assumed that the poet wants to approximate the rhythmic pattern to those established by tradition.
- 3) *Phonetic knowledge*. In Spanish, synaloepha is a natural phonetic phenomenon, while the dyalepha and dieresis sound artificial. Therefore, verses with dieresis or dyalepha are less likely to be a candidate for the scansion.

Therefore, the verse's scansion in Example 3 is performed as follows. Reading the whole poem,<sup>3</sup> we observe that the most frequent verse length is 11, thus Verses 3 a and 3 d are discarded. Then, the metric pattern of 3 b and 3 c are compared with the patterns from Table 1. We conclude that the verse from 3 b presents the characteristic accents 1.4.8.10 of the pattern *sáfico puro pleno* while pattern 1.4.7.10 from 3 c does not appear. Therefore, the correct scansion of the verse *Todas las tardes se muere un niño* is 1.4.8.10

As we have shown in this Section, the scansion of Spanish poetry is a challenging problem. In the next two Sections, we aim at automating the extraction of these metrical patterns of stressed and unstressed syllables with two different approaches.

#### IV. RANTANPLAN: SYLLABIFICATION AND SCANSION

The limitations mentioned above guided the design of Rantanplan [31], a modular syllabification and scansion system which is comprised of four modules that work together to perform scansion of both fixed-metre as well as mixed-metre poetry: PoS tagger, syllabification, stress assignment, and metrical adjustment.

##### A. MODULE 1: POS TAGGER

Rantanplan (see Algorithm 1) operates at the line level with a sequence of words. First, for each word in a line of the verse, the part of speech information is extracted since, in Spanish, some words are stressed depending on their function in the sentence.

<sup>3</sup><https://www.poesi.as/index239.htm>

##### Algorithm 1 Scansion Procedure

---

**Input:** A sequence  $\mathcal{W}$  of words  $\langle w_1, w_2, \dots, w_n \rangle$   
**Input:** A value *length* for the metrical length expected (optional)  
**Output:** A sequence  $\langle s_1, s_2, \dots, s_{\mathcal{L}} \rangle$  of booleans expressing the metrical pattern

**for**  $w_i \in \mathcal{W}$  **do**  
    $tag_i \leftarrow \text{pos}(w_i)$   
    $syllables_i \leftarrow \text{syllabify}(w_i)$   
    $stresses_i \leftarrow \text{stress}(syllables_i, tag_i)$   
 $groups \leftarrow \text{phonological}(syllables, stresses)$   
 $pattern \leftarrow \text{transform}(groups)$   
**if** *length* **then**  
   **while**  $|pattern| < length$  **do**  
       $g \leftarrow \text{generate\_phonological}(\mathcal{W})$   
       $pattern \leftarrow \text{transform}(g)$   
**return** *pattern*

---

For this purpose, Rantanplan was built on top of the industrial-strength NLP framework spaCy, which has notable speed performance [32]. However, spaCy splits most affixes thus causing some failures in the tags it assigns on prediction. To solve this limitation and to ensure clitics<sup>4</sup> were handled properly, Freeling's affixes rules via a custom-built pipeline for spaCy were integrated. The resulting package, spacy-affixes [33], fixes this bug and can be plugged into a regular spaCy pipeline loading one of the statistical models for Spanish. In Rantanplan, only suffixes on verbs are enabled in spacy-affixes to guarantee clitics are handled adequately, and PoS tags are assigned correctly.

##### B. MODULE 2: SYLLABIFICATION

With this information, the process of syllabification starts. Although this process follows a rule-based algorithm inspired by Mestre [34], Caparrós [23] and Tomás [22], and is based on the official rules for syllabification [28], some challenging situations were considered to boost the performance of the system. The first of these cases was compounds words and words with an 'h' in between. As an example of the former let us take the word 'reutilizar'. Although intuitively it may seem that the prefix 're-' should be separated from the rest of the word, the Fundéu<sup>5</sup> recommends not to do it this way, splitting instead as 'reu-ti-li-zar'. Similarly, the 'h' in a middle position does not split diphthongs, so 'desahijar' would be syllabified as 'de-sahi-jar', which might feel odd at a first pass, but it agrees with the pronunciation of the word. Another situation was related to the 'tl' group. The special feature of this group is that its syllabification changes according to the territory [35] but since most of the Spanish speakers around

<sup>4</sup>Syntactically independent but phonologically dependent morphemes that appear together in a word, e.g., in 'cógemelo', both 'me' and 'lo' are pronouns written together with the verb 'coge'

<sup>5</sup>The Fundéu is a foundation created from the *Department of Urgent Spanish* of the EFE Agency. See <https://twitter.com/Fundeu/status/1182226555457724416>

the world do not separate it, Rantanplan either does not split it. Moreover, diaereses are also included as part of the alternative syllabification exceptions (e.g. word ‘suave’, which poets tend to apply diaeresis to thus resulting in ‘sua-ve’ instead of the default split as ‘su-a-ve’) [36]. Hence, Rantanplan relies on a list of words with alternative syllabifications compiled from Ríos Mestre’s work. These alternatives are only taken into account by the metrical adjustment module.

### C. MODULE 3: STRESS ASSIGNMENT

Once syllables and part of speech of a word are extracted, stress can be assigned. The assignment of stress follows very closely the rules defined in [28] (e.g. oxytone, paroxytone, and proparoxytone syllables), adding exceptions for certain parts of speech, consonant groups, and words that are usually stressed but are not for metrical reasons. The sequence of phonological groups that will be used to extract the metrical pattern is calculated by applying all possible synaereses and synaloephas to the list of syllables of words per line, and propagating the stress information when needed. For example, the words ‘me ama’ are split into the syllables ‘me-a-ma’, and after applying synaloepha the resulting phonological groups, ‘**me**a-ma’, keep the stress in place. Word ends are also marked since they are needed to adjust the length of the metrical pattern according to the position of the stress of the last word. Phonological groups are then transformed into a metrical pattern representation and returned if the verse’s expected metrical length is not known beforehand.

### D. MODULE 4: METRICAL ADJUSTMENT

However, there are situations where the expected metrical length is known, such as processing a corpus of sonnets which tend to be hendecasyllables. In cases like this, verses with applied synaloephas or synaereses but a metrical length lower than the expected would trigger the adjustment module. In example 4, the expected metrical length is 11, but our system returns 9, thus triggering the metrical adjustment module.

- (4) *loor a mi autor, y al que leyere*  
*loor-a-miau-tor-y-al-que-le-ye-re*  
 metrical pattern: 1.4.8 – 9 syllables < 11  
 (Juan de Timoneda)

This means that  $11 - 9 = 2$  of the applied synaloephas and synaereses must be undone until both lengths match. The metrical adjustment module tries every possible metrical pattern combining synaereses, synaloephas, and alternative syllabifications. Priority is given to keep the synaloephas since they are rarely broken, and synaeresis are usually undone. The same happens for the alternative syllabifications, which deals with diaeresis and adds more combinations to check for. A special case adding possibilities to the search space is handling synaloephas between words with an initial ‘h’ and vowel ending words. Up to the 16th Century, the initial ‘h’ in words was aspirated instead of silent. This depends on the etymology of some words. For example, in the verse ‘*cubra*

*de nieve la hermosa cubre*’ (see example 5) there should not be synaloepha between ‘la’ and ‘hermosa’ since ‘hermosa’ evolved from the Latin ‘fermosa’ and as such a synaloepha was not possible at all. To this day, this remains an option to the author, who can decide whether or not to apply a synaloepha for cases like these.

- (5) *cubra de nieve la hermosa cubre*  
*cu-bra-de-nie-ve-la-her-mo-sa-cum-bre*  
 metrical pattern: 1.4.8.10 – 11 syllables  
 (Garcilaso de la Vega)

For each attempt, a new metrical pattern and length are calculated and checked against the expected metrical length. If no match is found, the last pattern calculated is returned. For the verse, in example 5, the generated possible metrical patterns are shown in example 6. Pattern 6.a, with no synaeresis and one synaloepha between ‘y’ and ‘al’ would be generated first and returned afterwards. Since the metrical pattern has the correct length, it is returned as such, and the generation stops, saving the time it takes to generate any other possible pattern. This is also a limitation of our approach since more than one correct metrical pattern can be generated that matches the desired length.

- (6) *loor a mi autor, y al que leyere*  
 a) *lo-or-a-mi-au-tor-y-al-que-le-ye-re*  
 metrical pattern: 2.6.10 – 11 syllables  
 b) *lo-or-a-miau-tor-y-al-que-le-ye-re*  
 metrical pattern: 2.5.10 – 11 syllables  
 c) *loor-a-mi-au-tor-y-al-que-le-ye-re*  
 metrical pattern: 1.5.10 – 11 syllables

### V. JUMPER: SCANSION WITHOUT SYLLABIFICATION

Jumper performs scansion of both fixed-metre and mixed-metre poetry based on three main assumptions:

- 1) The problem of counting syllables is different in nature from syllabification. The vowel is the nucleus of the syllable in Spanish and identifies its unit [37, 8.1a]. Thus, we can develop an algorithm for computing syllables and assigning stresses, which could significantly improve efficiency without using external libraries.
- 2) When the system is faced with an ambiguous verse, its metrical pattern’s possible realizations are finite and based on tradition. Therefore, we can resolve ambiguities by considering its approximation to the verse type’s natural patterns.
- 3) Scansion systems must consider the semiverses compensation of verses above eleven syllables analysis of mixed-metre poetry.

Based on these assumptions, we develop a system that implements a new way to count syllables, consider the compensation of the semiverses and resolve metrical ambiguities without losing information about the system’s decision. Jumper’s output gives the number of syllables, the metrical pattern, its name, and the labeled verse that led the system to extract the pattern.

The method comprises three modules: word, verse and poem analysis.

### A. MODULE 1: WORD LEVEL

For counting syllables efficiently, it is enough to count the vowels of a word, taking diphthongs into account. Given the contiguous occurrence of two vowels, the algorithm checks whether it is a diphthong not counting that following vowel as a syllable.

Also, the same procedure allows the method to assign the stress. The system assumes that the word follows the Spanish graphic accentuation rules [28, 3.4]. Thus, while counting vowels, the position in which a graphically accented vowel has been found is saved. After counting, if the stress has not been found, Spanish orthographic rules are applied.

### B. MODULE 2: VERSE LEVEL

This module is the method core. It consists of three submodules. The first module counts syllables and assigns the stresses to the whole verse. The second computes verse compensation for lines longer than eleven syllables. The third module resolves metric ambiguities.

#### 1) SUBMODULE 1: COUNTING SYLLABLES AND STRESSES IN A LINE OF VERSE

The first submodule works as follows. Verses are transformed into a list of words. For each word, the number of syllables, the position of stress in those syllables and the compensation factor are determined. The number of syllables of the word is added to the total number of syllables of the line, taking into account, if any, the synaloepha with the following word. The stress position is calculated and added to the list of the metrical pattern. If the word is unstressed, it is not added to the stresses list. In the first submodule, the verse's length is calculated considering all the synaloepha because it is the most natural rhetorical device. Subsequently, if it would not have to be considered for correct scansion, it is settled in the ambiguities module.

#### 2) SUBMODULE 2: SEMIVERSE COMPENSATION

At scansion beginning, verse length is unknown. If, after the computation of the first submodule, the line length results longer than eleven syllables, the metrical pattern's extraction is again performed, now considering the compound verse's metrical phenomena: the semiverse compensation and the dyalepha at the boundary between semiverses. For calculating the compensation, the algorithm must find the word that lies on the border between semiverses. A word is on the border of a semiverse when the line length, added to the number of syllables of the word and its compensation factor, results in the expected length of the semiverse. The expected semiverse length is calculated by dividing by 2 the syllabic length given by Module 1 because most compound lines have symmetrical semiverses.

### 3) SUBMODULE 3: AMBIGUITY RESOLUTION

The ambiguity resolution submodule accomplishes two tasks. The first one consists of detecting the possible metric ambiguities derived from the rhetorical devices and generating the candidate metric patterns. The second task consists of choosing, among the possible candidates, the best one.

The upper module of poem analysis indicates the activation of the ambiguity detection submodule. Once indicated, for efficiency, ambiguity detection (first task) is performed at the same time as the computation of Submodule 1. Once all the ambiguous candidates have been stored, they are combined: for example, a line can have both a dyalepha and a synaeresis. All candidates are labeled with their rhetorical devices to preserve the ambiguity resolution submodule's decision.

In the second task, the best candidate is chosen. As stated in the Problem definition section, the verses' metrical patterns are finite and their positions are close to the types of verses established by tradition. Therefore, the proximity of the extracted metrical pattern is compared with each of the types established in the literary precepts. Based on this proximity, Jumper rates each candidate, indicating the overlap degree with the most similar patterns. The candidate with the highest overlap degree is chosen.

### C. MODULE 3: POEM LEVEL

The poem is converted into a list of verses. Scansion is performed for each verse by Module 2; the ambiguity resolution submodule is disabled during this first computation.

Once all metrical patterns have been extracted, the frequency of the number of syllables of all the verses is computed. Two cases can occur:

- 1) If there is only one frequent length, it is concluded that it is a fixed-metre corpus.
- 2) If there are frequent measures is greater than one (for example, in contemporary poetry, poems with lines of 7, 11, and 14 syllables are common), the system concludes that it is a mixed-metre poem.

Finally, the algorithm recomputes the verses that do not match the length-frequency list, activating the ambiguity resolution submodule.

Once all ambiguities have been resolved, the scansion is finished.

## VI. EVALUATION

In recent years, the Corpus of Spanish Golden-Age Sonnets (henceforth: ADSO) by [36] has been used as the baseline for fixed-metre poetry scansion. It is a set of 10268 verses manually annotated with its metrical pattern. In his original work, Navarro-Colorado used a subset of 100 poems to evaluate its system for which an inter-annotator agreement of 96% was reported. Not all poems included in this subset are included in the larger publicly available ADSO corpus. Although smaller in size and virtually subsumed<sup>6</sup> by the larger corpus, it is the

<sup>6</sup>1387 out of 1400 ADSO100 verses are included in the large 10000 verses corpus.

**TABLE 2.** Accuracy and time on Navarro-Colorado's fixed-metre 10,000 verses corpus.

Method	Accuracy	Time
Gervás <sup>8</sup>	67.56	N/A
Navarro-Colorado	90.89	16,787s
Rantanplan	93.36	28s <sup>9</sup>
Jumper	94.68	3s

**TABLE 3.** Accuracy and time on Carvajal's mixed-metre 4,300 verses corpus.

Method	Accuracy	Time
Navarro-Colorado	49.38	7,484s
Rantanplan	78.67	10s
Jumper	81.73	1s

only subset for which inter-annotator agreement is reported, and therefore we are reporting on both.<sup>7</sup>

For evaluating mixed-metre poetry, we use our own corpus of over 4,300 verses obtained from Carvajal's annotated anthology [39]. It is composed of 4378 manually labeled verses of different lengths.

Systems are evaluated using their most recent versions (as described in this paper) according to their accuracy. The score is 1 if all and only the annotated stresses of a verse have been correctly extracted and 0 otherwise. System performance is measured as the accuracy rate over the total number of verses in the corpus. Accuracy is reported as a percentage with two significant figures. Time is given in seconds rounded to the closer integer. All evaluations were run on a computer with an Intel® Core i7-8550U CPU @ 1.80GHz and 16GiB of DDR4 RAM memory.

## A. EXPERIMENTAL RESEARCH

Tables 2 and 3 show accuracy and processing time for each of the methods on the 10268 verses of the fixed-metre corpus and the 4378 verses of mixed-metre corpus. Rantanplan scores 93.36 of per-line accuracy on fixed-metre and 78.67 on mixed-metre in 28 and 10 seconds, respectively after warmup (15 seconds). Jumper obtains 94.68 in 3 seconds on fixed-metre and 81.73 in 1 second on mixed-metre.

In the subset of 100 poems by Navarro-Colorado (see 4), for which inter-annotator agreement is known (96%), the accuracy of Jumper is close to inter-annotator agreement (94.97) and Rantanplan even exceeds it (96.94), leaving little margin for improvement on this subset. Rantanplan is almost three orders of magnitude faster than Navarro-Colorado. Jumper, in turn, is almost 10 times faster than Rantanplan. The explanation is that Jumper does neither require the compilation of numerous regular expressions for syllabification, nor the PoS-Tagging models for stress assignment.

<sup>7</sup>We used the tool Averell [38] to download and process both corpora

<sup>8</sup>Only 1,291 verses of the 1,400 verses corpus were evaluated by Gervás' method.

<sup>9</sup>We do not consider Rantanplan warmup time (15 seconds), since it is executed only once at the beginning of the scansion.

**TABLE 4.** Accuracy and time on Navarro-Colorado's fixed-metre 1,400 verses corpus.

Method	Accuracy	Time
Gervás <sup>10</sup>	70.87	N/A
Navarro-Colorado	94.44	2,356s
Rantanplan	96.94	3s
Jumper	94.87	0.31s

**TABLE 5.** Contingency tables of correct and incorrect assignments of line stress for the three corpora in this study.

		Jumper		
		Rantanplan	Incorrect	Correct
ADSO100	Incorrect		24	19
	Correct		48	1,313
ADSO	Incorrect		358	324
	Correct		188	9,399
Carvajal	Incorrect		697	237
	Correct		103	3,341

On the full fixed-metre poetry corpus, Jumper performs consistently better with a 2% relative improvement over Rantanplan. On the mixed-metre corpus, the accuracy of Jumper gives a 3% improvement with respect to Rantanplan. In terms of efficiency, Jumper runs between 11 and 9 times faster than Rantanplan. These results confirm that skipping syllabification and statistical language models is more efficient, as long as the richer output provided by Rantanplan is not needed. The speed achieved by Jumper allows exhaustive disambiguation, which also improves accuracy on both corpora.

We have applied two statistical tests to find out if the differences between Rantanplan and Jumper are statistically significant. First, we have applied McNemar's test [40], following [41], on the contingency tables displayed in Table 5. For the two larger datasets, Carvajal's and Navarro-Colorado 10,000 verses corpora (ADSO), the corresponding McNemar's approximated tests with continuity correction [42] indicate that the differences in the systems' errors are statistically significant:  $\chi^2(1) = 35.59, p < 0.001$  and  $\chi^2(1) = 52.03, p < 0.001$ , respectively, with Jumper performing better for both corpora. For the reduced version of Navarro-Colorado's corpus (ADSO100), we replace the approximated test with the exact test, because two components of the contingency table are too small ( $<25$ ) to apply the approximated version [40]. The results indicate that the difference between both systems is statistically significant ( $\chi^2(1) = 12.55, p < 0.001$ ), with Rantanplan performing better in this case.

We have also tested statistical significance by computing systems' performance on 40 random samples drawn from each of the two largest corpora (ADSO and Carvajal)<sup>11</sup>

<sup>11</sup>We apply this test only to the two larger corpora because the data samples of the reduced version of ADSO are not large enough to be statistically representative.

without replacement. We check the Student's t-test [43] requirements and, given that they are fulfilled, we use the t-test on the results.

We perform Shapiro and Wilk [44] and Levene's [45] tests to check normality and homoscedasticity, respectively (requirements to apply the t-test). Shapiro's statistics and p-values are  $s = 0.958$ ,  $p = 0.144$  for ADSO and  $s = 0.955$ ,  $p = 0.115$  for Carvajal. Levene's test results are  $s = 1.866$ ,  $p = 0.179$  for ADSO and  $s = 0.001$ ,  $p = 0.974$  for Carvajal. In all these tests,  $p > 0.05$ , which indicates that the error distributions are normally distributed and present homogeneity of variance. Therefore, we can perform the t-test. The results indicate that, in both cases, there is a statistically significant difference between systems:  $t = 4.271$ ,  $p < 0.001$  and  $t = 3.921$ ,  $p < 0.001$  for ADSO and Carvajal corpora, respectively, with Jumper performing better in both cases.

In the following section, we carry out an in-depth qualitative error analysis in order to reach a better understanding of the (comparative) strengths and weaknesses of both systems.

## B. ERROR ANALYSIS

This section delves into the circumstances in which Rantanplan and Jumper fail to classify metric patterns. Rantanplan misclassifies 682 out of 10268 (6.64%) in the fixed-metre corpus. Jumper, 546 (5.32%). On the more challenging mixed-metre corpus, Rantanplan and Jumper misclassify 934 and 800 out of 4378 (21.33% and 18.27% error rates) verses, respectively.

In our analysis, we distinguish three types of errors:

- 1) Errors in common with identical outputs: Rantanplan and Jumper have extracted the same pattern, but it does not match the manual annotation.
- 2) Errors in common with different outputs: those patterns in which both systems fail, but each one has extracted different patterns.
- 3) Errors that are exclusive to each system: those patterns that Rantanplan has classified wrong, but Jumper has classified right and vice versa.

We manually analyzed a sample of 99 verses for each dataset, consisting of 33 random samples of each error type.

### 1) ERRORS ON FIXED-METRE CORPUS

Figure 1a shows the distribution of the three types of errors on the fixed-metre corpus. Both systems extract the same (incorrect) pattern for 2.68% of the verses examined. During the manual inspection, we detect that 6 out of 33 errors in this subset are attributable to the systems alone, and 4 are of dubious origin. However, a majority of cases (19 out of 33) are not actually system errors, but incorrect manual annotations. Inspecting the cases where both systems give the same wrong answer, therefore, is a good method to clean up manual annotation errors.

Errors in common with different extracted patterns account for 0.81% of the verses. In this case, 14 out of the 33 cases manually inspected are due to system failures, and 5 are

annotation errors. Although it is a marginal percentage, the manual observation is interesting since 13 of 33 are errors whose origin is ambiguous. The correct scansion of these verses is also difficult for humans, and both systems offer valid solutions. Example 7 shows a verse of this kind of error. It presents four possible synaloephas; its high ambiguity leads that all three scansions (the manual annotation and the ones proposed by Rantanplan and Jumper) are actually possible.

- (7) *iAy dura, ay importuna, ay triste ausencia!*  
 a)  
*Ay-du-ra-ayim-por-tu-na-ay-tris-teau-sen-cia!*  
 1.2.4.6.8.10 – 11 syllables  
 (annotation)  
 b)  
*iAy-du-ra-ayim-por-tu-na-ay-tris-teau-sen-cia!*  
 2.6.8.10 – 11 syllables  
 (Rantanplan)  
 b)  
*iAy-du-ra-ay-im-por-tu-na-ay-tris-teau-sen-cia!*  
 metrical pattern: 2.4.7.8.10 – 11 syllables  
 (Jumper)

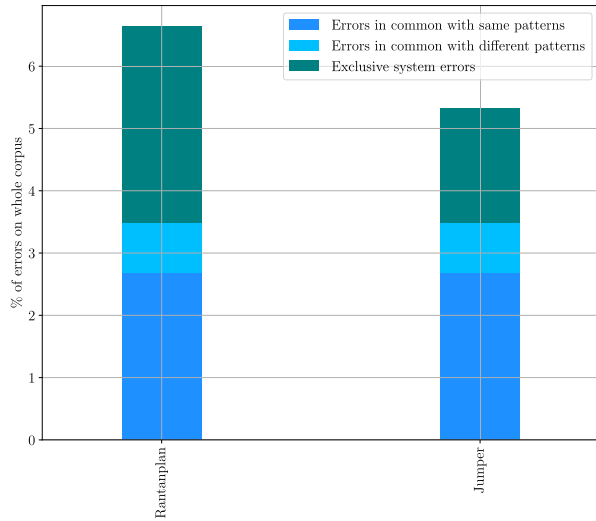
Finally, there are 3% errors exclusive to Rantanplan and 1.8% to Jumper. In the manual analysis, all of them are true system errors. Both systems incur in the same type of mistakes:

- When a verse presents combined rhetorical devices, occasionally, systems do not disambiguate correctly. Example 8 shows a verse with two possible scansions derived from the ambiguity of the dyaloepha in *italianas*; it can be read as *i-ta-lia-nas* or *i-ta-li-a-nas*. By applying the disambiguation rules explained in the Problem definition section (III), we conclude that the Option B is the correct analysis of the verse.

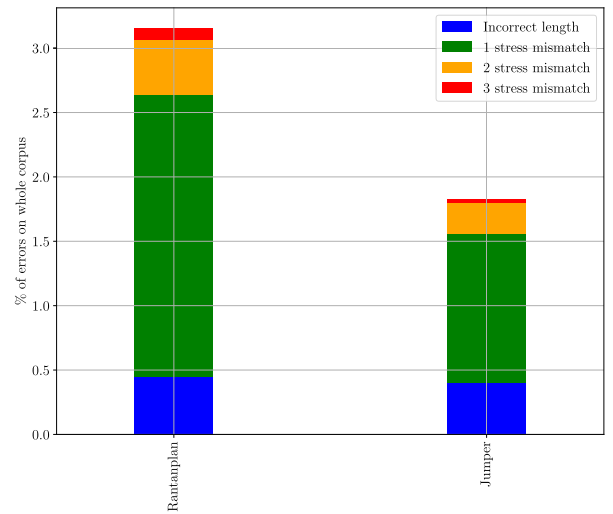
- (8) *Musas italianas y latinas*  
 a) *Mu-sas-i-tali-a-nas-y-la-ti-nas*  
 1.6.10 – 11 syllables  
 Correct (annotation and Jumper)  
 b) *Mu-sas-i-ta-lia-nas-y-la-ti-nas*  
 1.5.9 – 10 syllables  
 Incorrect (Rantanplan)

- The methods sometimes consider unstressed a syllable that must be stressed given its position in the pattern and vice versa. Example 9 shows a verse in which the word *oh* is not stressed because the adjacent stress of *dueño* overrides it.

- (9) *iOh dueño sin piedad, que tal ordenas!*  
 a) *iOh-due-ño-sin-pie-dad-que-tal-or-de-nas!*  
 2.6.8.10 – 11 syllables  
 Correct (annotation and Rantanplan)  
 b) *iOh-due-ño-sin-pie-dad-que-tal-or-de-nas!*  
 1.2.6.8.10 – 11 syllables  
 Incorrect (Jumper)



(a) Distribution of the three types of errors analyzed for each system.



(b) Analysis of the systems' exclusive errors indicating how close the methods have been at classifying correctly the pattern.

FIGURE 1. Errors on fixed-metre corpus.

We now perform a more detailed quantitative analysis on this third subset of errors, because in this case all failures are due to systems decisions. We want to know how close the methods have been at classifying the pattern correctly. We consider that a system failed at correctly classifying a pattern if it was unable to extract the verse length.<sup>12</sup> On the contrary, we consider that a method was close to correctly classifying a well when it misassigned only one stress. Out of the 3% of errors made by Rantanplan, it only failed to assign one stress in 2.19% of the cases. Similarly, Jumper only failed to assign one stress in 1.15% of 1.8% incorrectly extracted patterns (Figure 1b).

## 2) ERRORS ON MIXED-METRE CORPUS

We performed the same error analysis on the mixed-metre corpus. Figure 2a shows their distribution. For 13.7% of the verses, both systems extracted the same pattern and this pattern did not match the annotation in the corpus. But only 3 out of the 33 errors analyzed manually were actually system errors, and 2 additional cases were ambiguous. The vast majority of the discrepancies, 28 out of 33, were actually manual annotation errors. Again, we can improve the manual annotation in the corpus efficiently just by checking these cases. This dataset, in particular, seems to have a higher number of manual errors.

For 2.22% of the verses, the systems extracted different patterns, and none of them matched the manual annotation. A total of 5 out of the 33 cases in this situation

<sup>12</sup>The Spanish versification is based on the regularity of the verse's last mandatory stress. This last stress is the one on which the previous stresses are set and, therefore, it gives rise to the stress structure of the verse type. Thus, if the last stress is not correctly extracted, we consider that the scansion has failed.

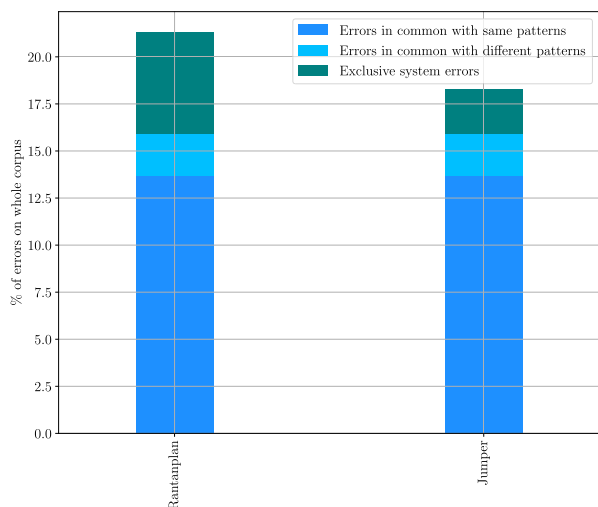
were system failures, and 13 were manual annotation errors. For the remaining 14 cases, the attribution was ambiguous. Example 10 presents a verse that appears in this kind of errors but the three solutions are valid depending on the reader's diction.

- (10) *enviada en tu voz, por mí asumida*  
 a) *en-vi-a-da-en-tu-voz-por-mía-su-mi-da*  
 3.6.8.10 – 11 syllables  
 (annotation)  
 b) *en-via-da-en-tu-voz-por-mí-a-su-mi-da*  
 2.5.7-10 – 11 syllables  
 (Rantanplan)  
 c) *en-via-da-en-tu-voz,-por-mía-su-mi-da*  
 2.6.8.10 – 11 syllables  
 (Jumper)

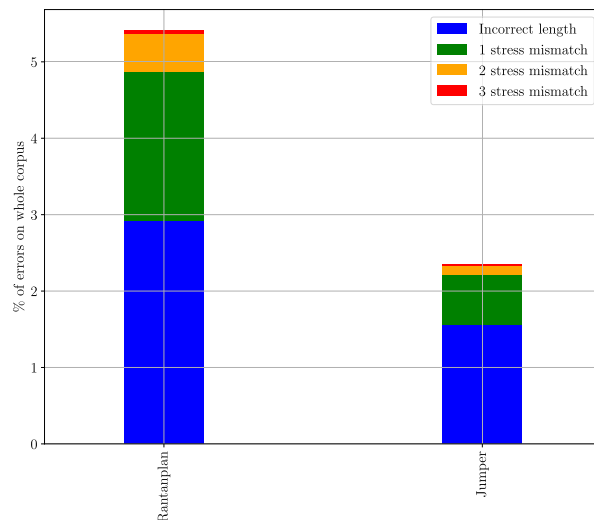
Finally, there was a 5.41% of errors exclusive to Rantanplan and a 2.35% exclusive to Jumper. Figure 2b shows the errors caused by incorrect length assignment and stress mismatches. The errors made by the two systems are due to the same reasons as in the fixed-metre corpus. Example 11 shows an incorrect disambiguation.

- (11) *que se hunda en la nada*  
 a) *que-sehun-da-en-la-na-da*  
 2.6 – 7 syllables  
 Correct (annotation and Rantanplan)  
 b) *que-se-hun-da-en-la-na-da*  
 3.6 – 7 syllables  
 Incorrect (Jumper)

Example 12 shows an incorrect consideration as unstressed the word *oh*, which must be considered stressed given its position in the pattern.



(a) Distribution of the three types of errors analyzed for each system.



(b) Analysis of the systems' exclusive errors indicating how close the methods have been at classifying correctly the pattern.

FIGURE 2. Errors on mixed-metre corpus.

- (12) *oh, corazón de otoño tranquilamente abierto*  
 a) *oh-co-ra-zón-deo-to-ño-tran-qui-la-men-tea-bier-to*  
 1.4.6.9.11.13 – 14 syllables  
 Correct (annotation and Jumper)  
 b) *oh-co-ra-zón-deo-to-ño-tran-qui-la-men-tea-bier-to*  
 4.6.9.11.13 – 14 syllables  
 Incorrect (Rantanplan)

Both systems make these two error types.

Also, the mixed-metre corpus contains verses of more than 11 syllables, therefore susceptible to the phenomena between semiverses. We analyzed the impact of handling the semi-verses since it is one of the differences between Rantanplan and Jumper. Of the 4378 verses in the corpus, 832 (19%) were compound verses, of which only 43 (1%) manifest the compensation phenomena. This 1% alone is responsible for the 3.8% discrepancy between Rantanplan and Jumper. Example 13 shows the scansion of this kind of verse from Rantanplan and Jumper output. The word *cálido* is proparoxytone and it is located in the limit of the semiverse. Thus, one syllable less is counted. Adding the correct treatment of semiverses in Rantanplan should therefore boost its accuracy.

- (13) *en los latidos cálidos del vientre y de las sienes.*  
 a) *en-los-la-ti-dos-cá-li-dos-del-vien-trey-de-las-sie-nes.*  
 4.6.9.13 – 14 syllables  
 Correct (annotation and Jumper)  
 b) *en-los-la-ti-dos-cá-li-dos-del-vien-trey-de-las-sie-nes*  
 4.6.10.14 – 15 syllables  
 Incorrect (Rantanplan)

## VII. RANTANPLAN VS JUMPER: DISCUSSION

The results of our experimentation show that both Rantanplan and Jumper bring substantial improvements with respect the state of the art. Comparing both systems, Jumper is substantially faster and slightly more accurate, while Rantanplan provides richer linguistic information. Remarkably, using both systems on the same corpora has proved to be particularly useful in our experimentation: when both systems make the same mistake, it is a reliable signal that the manual annotation might be erroneous. Therefore, used in conjunction they can be a useful tool for corpus clean-up.

The different approaches behind both algorithms make them complementary, and the best suited for each scenario depends on the user needs. Rantanplan is a scansion tool conceived to assist researchers on the metrical analysis of a work (i.e., a poem) and to be used as a source of information for populating a new ontology devoted to poetic works proposed by the POSTDATA ERC project [46]. This ontology is a three-layer encapsulated ontology aligned with FRBRoo foundational ontology [47] that models the primary information of a poem from the abstract concept of poetic work (i.e. FRBRoo work class), through the textual information of poem (i.e. expression FRBRoo class) where all the poem metrical information is included with a fine degree of granularity [48]. All the relations among stanzas, lines and syllables and the information of the rhetoric devices (i.e. synaloepha, synaeresis) are calculated and stored. The last layer adds all the bibliographic information (i.e. manifestation FRBRoo class). Currently, Rantanplan is being used for the population of the ontology by preprocessing different free repertories and databases while providing a very accurate and rich analysis [49]. Rantanplan stores further useful information to deal with the uncertainty on different abstraction levels in the analysis of a poem. This is possible because it calculates not

only the most likely verse length but also the verse length interval (depending to the use or not of rhetorical devices). Storing this uncertainty enables a new kind of analysis in poetry such as textual criticism or literary criticism. Finally, the fine-grained information extracted by Rantanplan allows for other uses, such as teaching poetry or as an analysis tool for more complex tasks such as lyrics classification (e.g. LyrAICs [50] or as a tool to develop on its top new digital humanities infrastructures, [51]). Rantanplan is integrated in the larger PoetryLab, a web-based application and API that openly exposes the services and technologies developed under the POSTDATA Project. In terms of performance for the scansion task, Rantanplan improves all other systems (except Jumper) in terms of accuracy and efficiency. Compared to Jumper, its one-solution-fits-all approach negatively impacts its running time.

Jumper, on the other hand, is a specific-purpose algorithm focused on fast and accurate metrical pattern extraction. As a rule-based method, Jumper stores useful information to preserve the system decision: all ambiguous candidates are labeled with their rhetorical devices (i.e. synaloepha, synaeresis); they are rated with a metric that indicates a degree of self-assessed confidence. Its accuracy is even better than Rantanplan's state of the art performance and, most remarkably, is nearly ten times faster in running time. Jumper's algorithm has been deployed in JumperApp, a real-time poetry analysis application designed to assist poets, philologists or translators. It is currently a standalone development – with almost no dependence on external libraries and modules–, which makes its use straightforward for practitioners. On the other hand, its potential as a module for other tasks is yet to be explored, and it is less suited for pedagogical purposes, where explicit syllabification can help learners understand and perform metric analysis.

## VIII. CONCLUSION

In this work, we have presented two systems that solve the scansion task for the Spanish poetry analysis: Rantanplan, which rely on a NLP pipeline that includes syllabification and PoS-tagging, and Jumper, that solves the problem without syllabification.

We have evaluated the systems on a fixed-metre and a mixed-metre corpora. The results indicate that the syllabification-free algorithm is a more efficient approach, as long as the richer output provided by Rantanplan is not needed. Since neither the compilation of numerous regular expressions for syllabification nor the PoS-tagging models for stress assignment are required, the running time of Jumper is notably reduced. However, Rantanplan's usefulness for certain primary use cases is not reflected in the quantitative results. The output produced by Rantanplan is machine-readable, interoperable, and ready to be ingested into a linked open data triple store compliant with the POSTDATA Project network of ontologies. Overall, both systems are more accurate and more efficient than previous methods.

In our error analysis, we have detected two primary sources of error: when a verse presents combined rhetorical devices, occasionally, none of the systems disambiguate correctly. The methods sometimes consider unstressed a syllable that must be stressed given its position in the pattern. These error sources indicate the parts of the systems that can be improved in future versions.

Our error analysis has also revealed a positive side effect of processing the data with both systems. In those cases where Jumper and Rantanplan differ from the corpus, it is usually due to a wrong annotation. Therefore, the use of both tools is an efficient method to clean-up manually annotated corpus. Finally, we observe that for errors with different system patterns, both systems obtain valid alternative solutions. This highlights that the automatic scansion task will never be perfect, since it is related to the human creativity and human realization of a poem.

## APPENDIX CODE AND REPRODUCIBILITY

The code of Rantanplan is available for download at <https://github.com/linhd-postdata/rantanplan>, and the evaluation code at <https://github.com/linhd-postdata/rantanplan-evaluation>. Tables in this paper can be reproduced using the notebook: <https://colab.research.google.com/drive/1t-FxbA7aurJTgCoFUBFYzDmHHImX0k4m?usp=sharing>. JumperApp binaries and code are available for download at <https://github.com/grmarco/jumper>. Finally, a working demo of the PoetryLab UI is accessible at <http://postdata.uned.es/poetrylab/>, while its RDF-ready OpenAPI is hosted at <http://postdata.uned.es:5000/ui/>

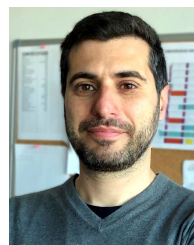
## REFERENCES

- [1] J. D. Caparrós, *Diccionario de Métrica Española*. Chennai, India: Alianza, 2001.
- [2] J. de la Rosa, Á. Pérez, L. Hernández, S. Ros, and E. González-Blanco, "Rantanplan, fast and accurate syllabification and scansion of Spanish poetry," *Procesamiento Lenguaje Natural*, vol. 65, pp. 83–90, Sep. 2020.
- [3] G. Marco Remón and J. Gonzalo, "Automatic scansion of Spanish poetry without syllabification," 2020, *arXiv:2012.12799*. [Online]. Available: <http://arxiv.org/abs/2012.12799>
- [4] E. González-Blanco, S. R. Mu noz, P. R. Fabo, M. L. D. I. Platas, H. B. Z. Sabel, A. C. C. Herraes, C. I. I. M. Cantó, and L. Ayciriex, "Poetry and digital humanities making interoperability possible in a divided world of digital poetry: POSTDATA project," in *Proc. EADH*, 2018, pp. 1–5.
- [5] C. O. Hartman, *Virtual Muse: Experiments in Computer Poetry*. Middletown, CT, USA: Wesleyan University Press, 1996.
- [6] H. M. Logan, "Computer analysis of sound and meter in poetry," *College Literature*, vol. 15, pp. 19–24, Jan. 1988.
- [7] M. Newman, "Poetry processing," *Byte Mag.*, vol. 11, no. 2, pp. 224–225, 1986.
- [8] M. De sexto, L. Hernandez, J. De la Rosa, A. Perez, S. Ros, and E. Gonzalez-Blanco, "Understanding poetry using natural language processing tools: A survey," in *Proc. UNED*, 2020, pp. 1–32.
- [9] S. Yokohama. (2019). *Digital Technologies for Exploring Prosody: A Brief Historical Overview*. Accessed: Feb. 28, 2021.
- [10] V. Beaudouin and F. Yvon, "The metrometer: A tool for analysing French verse," *Literary Linguistic Comput.*, vol. 11, no. 1, pp. 23–31, Apr. 1996.
- [11] C. O. Hartman. (2005). *The scandroid 1.1*. [Online]. Available: <http://oak.conncoll.edu/cohar/Programs.htm>

- [12] M. R. Plamondon, "Virtual verse analysis: Analysing patterns in poetry," *Literary Linguistic Comput.*, vol. 21, no. 1, pp. 127–141, Mar. 2006.
- [13] W. G. M. McAleese, "Improving scansion with syntax: An investigation into the effectiveness of a syntactic analysis of poetry by computer using phonological scansion theory," Dept. Comput. Fac. Math., Comput. Technol. Open Univ., Buckinghamshire, U.K., Tech. Rep. 2007/26, 2007. [Online]. Available: <http://computing-reports.open.ac.uk/2007/TR2007-26.pdf>
- [14] H. Anttila. (2010). *Prosodic*. Accessed: Feb. 2, 2021. [Online]. Available: <http://prosodic.stanford.edu>
- [15] R. Ibrahim and P. Plecháč, "Towards the automatic analysis of Czech verse," in *Proc. Formal Methods Poetics*, 2011, pp. 295–305.
- [16] K. Bobenhausen and B. Hammerich, "Métrique littéraire, métrique linguistique et métrique algorithmique de l'allemand mises en jeu dans le programme Metricalizer2," *Languages*, vol. 3, pp. 67–88, Jan. 2015.
- [17] K. Bobenhausen. (2017). *Metricalizer*. Accessed: Jan. 21, 2021. [Online]. Available: <http://www.meticalizer.de/en/>
- [18] D. Sisto, "The interaction between phonology and meter. Approaches to Romance and West-Germanic Renaissance meter," Ph.D. dissertation, Meertens Inst. Roy. Netherlands Acad. Arts Sci., Radboud Univ. Nijmegen, Amsterdam, The Netherlands, 2020.
- [19] E. C. Papakitsos, "Computerized scansion of ancient greek hexameter," *Literary Linguistic Comput.*, vol. 26, no. 1, pp. 57–69, Apr. 2011.
- [20] J. D. Caparrós, *Contribución a la Historia de las Teorías Métricas en Los Siglos*, vol. 92. Serrano, Madrid: CSIC-CSIC, 1975.
- [21] A. Bello, *Principios de la Ortolojía Métrica de la Lengua Castellana*. Los Angeles, CA, USA: La Opinión, 1859.
- [22] T. N. Tomás, "Métrica española," in *Reseña Históricay Descriptiva*, vol. 50. Dec. 1991, p. 556.
- [23] J. D. Caparrós, *Métrica Española*. Madrid, Spain: Síntesis, 1993.
- [24] P. Gervas, "A logic programming application for the analysis of Spanish verse," in *Proc. Int. Conf. Comput. Log.* New York, NY, USA: Springer, 2000, pp. 1330–1344.
- [25] B. Navarro-Colorado, "A metrical scansion system for fixed-metre Spanish poetry," *Digit. Scholarship Humanities*, vol. 33, no. 1, pp. 112–127, Apr. 2018.
- [26] L. Padró and E. Stanilovsky, "Freeling 3.0: Towards wider multilinguality," in *Proc. Int. Conf. Lang. Resour. Eval.*, 2012, pp. 1–7.
- [27] M. Agirrezabal, I. Alegria, and M. Hulden, "A comparison of feature-based and neural scansion of poetry," 2017, *arXiv:1711.00938*. [Online]. Available: <http://arxiv.org/abs/1711.00938>
- [28] R. A. E. RAE, *Ortografía de la Lengua Española*. Asturias, Spain: Espasa, 2010.
- [29] A. Quilis, *Métrica Española*. Barcelona, Spain: Ariel Barcelona, 1984.
- [30] P. J. Pou, *Métrica Española*. Madrid, Spain: Ediciones Cátedra, 2020.
- [31] P. E. project. (2021). *Rantanplan Description*. Accessed: Feb. 28, 2021. [Online]. Available: <https://github.com/linhd-postdata/rantanplan/>
- [32] M. Honnibal and I. Montani, "SpaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing," *Explosion Company Open Source*, Berlin, Germany, Tech. Rep. 10/2017, Jul. 2017.
- [33] P. E. project. (2021). *Spacy Affixes Description*. Accessed: Feb. 28, 2021. [Online]. Available: <https://github.com/linhd-postdata/spacy-affixes/>
- [34] A. R. Mestre, "La transcripción fonética automática del diccionario electrónico de formas simples flexivas del español: Un estudio fonológico en el léxico," Ph.D. dissertation, Dept. de Filología Espanyola, Univ. Autònoma de Barcelona, Barcelona, Spain, 1998.
- [35] Fundeu. (2021). *Fundeu Webpage*. Accessed: Feb. 2, 2021. [Online]. Available: <https://www.fundeu.es/consulta/at-lan-ti-co-o-a-tlan-ti-co-12213>
- [36] B. Navarro-Colorado, M. R. Lafoz, and N. Sánchez, "Metrical annotation of a large corpus of Spanish sonnets: Representation, scansion and evaluation," in *Proc. Int. Conf. Lang. Resour. Eval.*, 2016, pp. 4360–4364.
- [37] S. Rae, *Nueva Gramática de la Lengua Española Fonéticay fonologa*. Asturias, Spain": Espasa Libros, 2009, vol. 3.
- [38] A. Díaz, A. Pérez, and J. de la Rosa. (Sep. 2020). *Averell 1.1.0*. Accessed: Jan. 21, 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.4036061>
- [39] A. Carvajal, *Extravagante jerarquía: 1968-1981*, vol. 58. Santa Clara, CA, USA: Hiperión, 1983.
- [40] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, Jun. 1947.
- [41] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998.
- [42] A. L. Edwards, "Note on the 'correction for continuity' in testing the difference between correlated proportions," *Psychometrika*, vol. 13, no. 3, pp. 185–187, Sep. 1948.
- [43] A. Student, "The probable error of a mean," *Biometrika*, vol. 6, no. 1, pp. 1–25, Mar. 1908.
- [44] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, nos. 3–4, pp. 591–611, Dec. 1965.
- [45] H. Levene, "Robust tests for equality of variances," in *Contributions to Probability and Statistics. Essays in honor of Harold Hotelling*. Stanford, CA, USA: Stanford Univ. Press, 1961, pp. 279–292.
- [46] E. G.-B. García, M. Maniulescu, and S. Ros, "From syllables, lines and stanzas to linked open data: Standardization, interoperability and multi-lingual challenges for digital humanities," in *Proc. 4th Int. Conf. Technological Ecosystems for Enhancing Multiculturality*. New York, NY, USA: ACM, Nov. 2016, pp. 979–983.
- [47] Cidoc-Crm. (2021). *FRBoo Description*. Accessed: Jan. 21, 2021. [Online]. Available: <http://www.cidoc-crm.org/frbroo/>
- [48] M. D. Platas, S. Ros, E. Gonzalez-Blanco, O. Corcho, O. Khali, L. Hernandez, M. De sexto, J. De la Rosa, A. Perez, and A. Diez, "Description of postdata ontology version 1.0," in *Proc. 3rd Plotting Poetry Conf*. Berlin, Germany: De Gruyter, 2020, pp. 1–15.
- [49] P. E. project. (2021). *Averell Corpus Description*. Accessed: Feb. 2, 2021. [Online]. Available: <https://github.com/linhd-postdata/averell>
- [50] E. González-Blanco and S. Ros. (2021). *LyrAics Proof of Concept ERC Grant*. Accessed: Feb. 28, 2021. [Online]. Available: <https://erc.europa.eu/news-events/magazine/PoC-2020-second-round-highlights>
- [51] C. S. Group. (2021). *CLS Infra Horizon 2020 Infrastructural Project*. Accessed: Feb. 28, 2021. [Online]. Available: <https://bit.ly/3oMpJeP>



**GUILLERMO MARCO** received the B.S. degree in computer engineering and the master's degree in artificial intelligence from the Polytechnic University of Madrid. He is currently pursuing the Hispanic Philology degree. He is also a Hispanic Philology Poet. He is also a Predoctoral Fellow with the UNED Natural Language Processing and Information Retrieval Research Group. His first book won the Second Award of the Adonais Prize, one of the most important distinctions in Spanish poetry.



**JAVIER DE LA ROSA** received the Ph.D. degree in hispanic studies from the University of Western Ontario, with a focus on digital humanities, and the master's in artificial intelligence from the University of Seville. He has previously worked as a Research Engineer with the Center for Interdisciplinary Digital Research, Stanford University, and as the Technical Lead with the University of Western Ontario CulturePlex Laboratory for cultural complexity. He is currently a Postdoctoral

Fellow with UNED Digital Humanities Innovation Laboratory. His interest includes natural language processing applied to historical and literary text, to the analysis of networks of fine arts artifacts and the visual culture of the past.



**JULIO GONZALO** is currently a Full Professor with Universidad Nacional de Educación a Distancia (UNED), Spain, where he leads the Research Group in natural language processing and information retrieval. His publications cover many topics at the intersection of information retrieval and natural language processing. His current research interests include evaluation metrics and methodologies, semantic textual similarity, online reputation monitoring, and information access technologies for social media. He has Co-Chaired evaluation activities, such as RepLab (for Reputation Monitoring tasks), WePS (Web People Search), and iCLEF (interactive Cross-Language search and Question Answering). In this context, his research made a special emphasis on the formal assessment of evaluation metrics, which led to a Google Faculty Research Award (with Stefano Mizzaro and Enrique Amigó) in 2012. He has recently been appointed as the General Co-Chair for the ACM SIGIR 2022 Conference.



**SALVADOR ROS** (Senior Member, IEEE) received the M.Sc. degree in physics from Complutense University Madrid, Madrid, Spain, in 1991, with a focus on control and automatic systems. He has been the Director of Learning Technologies with UNED, for six years. He is currently a Senior Lecturer and the former Vice-Dean of technologies with the School of Computer Science, Spanish University for Distance Education, UNED, Madrid. He is also an Associate PI of the POSTDATA ERC Project. His research interest includes enhanced learning technologies for distance-learning scenarios. He received the Extraordinary Doctoral Award in UNED for his Ph.D. Dissertation and two special best paper awards.



**ELENA GONZÁLEZ-BLANCO** is currently pursuing the Ph.D. degree in Spanish philology. She was the Director and the Founder of LINHD. She is currently an Associate Professor of artificial intelligence applied to business with IE. She is also a General Manager of Europe at CoverWallet. She has been the Head of artificial intelligence product development at Minsait by Indra. She is also a Renowned International Researcher and a Principal Investigator of the H2020 ERC Project POSTDATA. She is also a member of the Executive Committee of the European Alliance for Digital Humanities, the Secretary of the International Alliance for Digital Humanities Organization, and the Advisory Board of the Clarin ERIC EU Research Infrastructure, among others. She is also recognized as one of the Top100 Female Leaders in Spain in 2016, 2017, and 2018, and awarded with the Julián Marías Prize 2017 for researchers under 40 years and the First National M.A. Prize in Spanish Studies and Classics. She has been one and three in the Choiseul Ranking “100 Economic Leaders for the Future of Spain” in 2018 and 2019, respectively, and one of the #40under40 in Insurance in Spain 2020.

...