The Automatic Quantitative Metrical Analysis of Spanish Poetry with Rantanplan: A Preliminary Approach

Laura Hernández-Lorenzo

Digital Humanities Innovation Lab, UNED, Spain
laura.hernandez@scc.uned.es
0000-0003-3489-2193

Mirella De Sisto

Digital Humanities Innovation Lab, UNED, Spain
mdesisto@scc.uned.es
0000-0002-0899-5976

Álvaro Pérez

Digital Humanities Innovation Lab, UNED, Spain
alvaro.perez@scc.uned.es
0000-0001-5897-1246

Javier de la Rosa

Digital Humanities Innovation Lab, UNED, Spain
versae@gmail.com
0000-0002-9143-5573

Salvador Ros

Digital Humanities Innovation Lab, UNED, Spain
sros@scc.uned.es
0000-0001-6330-4958

Elena González-Blanco

School of Human Sciences and Technology,
IE University, Spain
egonzalezblanco@faculty.ie.edu
0000-0002-0448-1812

Abstract

In this paper, we present a quantitative approach to Spanish poetry and versification based on the application of our own automatic metrical tool, Rantanplan, to the complete poetic works of four early modern Spanish poets. All of the poetry of these four representative authors—Garcilaso de la Vega (1503–1536), Fernando de Herrera (1534–1597), Luis de Góngora (1561–1627), and Lope de Vega (1562–1635)—was automatically processed and stress positions were extracted. Thanks to the development of a new stanza identification feature of Rantanplan, we were able to detect metrical structures as well. By completing a quantitative analysis of the stress positions, line lengths, and stanzas used by each author, we aim to model their complete metrical profiles.

1 Introduction

Applications of quantitative methodologies to poetry analysis have grown significantly in recent years driven by the emergence of automatic tools for verse

analysis. Although there have been fewer applications of these techniques to Spanish poetry, quantitative studies have already been carried out of the stress positions in sonnets and romances. We are, however, not aware of the existence of any computational studies that analyze stanza information extracted from the complete poetic works of an author.

In this article, we present a quantitative metrical analysis of four corpora of works by individual Golden Age Spanish poets. Using our own automatic scansion and syllabification tool, Rantanplan (De la Rosa et al. 2020a), which includes a new feature for stanza identification, we undertake the metrical annotation of the complete works of four canonical Spanish poets. These authors, Garcilaso de la Vega (1503–1536), Fernando de Herrera (1534–1597), Luis de Góngora (1561–1627), and Félix Lope de Vega (1562–1635), are selected based on the availability of their texts and their significance in this period. By studying and analyzing each poet's metrical choices related to stress positions, line lengths, and stanzas, we aim to model their complete metrical authorial profiles.

The paper is structured as follows: after this introduction, Section 2 contextualizes our research as part of an ongoing quantitative metrical investigation of poetry with a special focus on Spanish works. Section 3 then introduces our study, the dataset that we use, the process by which it was collected (Section 3.1) and the methodology that we employ (Section 3.2). Section 4 includes a quantitative metrical analysis which presents and discusses our results related to the stress positions (Section 4.1), line lengths (Section 4.2), and stanzas used by these authors (Section 4.3). Finally, our conclusions and proposals for future work are presented in Section 5. We also include an appendix of complementary materials.

2   Quantitative Approaches to (Spanish) Poetry Research

Since the emergence of what is today known as the digital humanities (Schreibman et al. 2004), traditional literary research has benefited from the application of new methods and techniques. In this regard, there has been a notable rise in quantitative approaches to the study of literary records, especially in light of theories of distant reading by Moretti (2003, 2013) and macroanalysis by Jockers (2013). As Burrows writes, "computer-assisted textual analysis can be of value in many different sorts of literary inquiry, helping to resolve some questions, to carry others forward, and to open entirely new ones" (Burrows 2004).

Turning to poetry, quantitative methodologies and Natural Language Processing techniques have been applied far less often than is the case with narrative texts. At first glance, the former genre has several limitations; among the most important are the generally shorter length of poetic texts and their subjective dimensions. Despite this, there has recently been significant scholarly attention to the automatic analysis of poetic works. We refer here in particular to several academic studies that apply these techniques, conferences and monographs dedicated to quantitative approaches to poetry studies (Plecháč et al. 2019), and a growing number of tools specifically developed for verse analysis. The latter include Natural Language Processing tools, which perform the automatic

scansion and syllabification of poems. However, to the best of our knowledge, there has not yet been any exploration of quantitative digital approaches based on a combination of line and stanza information.

Regarding literary traditions, most quantitative and automatic verse studies have addressed English language poetry, while fewer studies and tools are available for other European languages. In the case of Spanish poetry, several automatic scansion and syllabification tools have been developed since the first system built by Gervás (2000). Of these, Zeuscansion (Agirrezabal et al 2013), the ADSO Scansion system (Navarro-Colorado 2017), and Rantanplan (De la Rosa et al. 2020a) are the most recent. This rise in automatic tools has also encouraged quantitative metrical studies of Spanish poetry. These have mainly been carried out within ADSO[1] (Navarro-Colorado 2015, 2016) and the POSTDATA project[2] (Ruiz Fabo et al. 2020). Additional work has taken place within "Verse Rhythm in Golden Age Spanish Poetry: Lope de Vega and Luis de Góngora's Romances," a quantitative metrical project that does not use digital methods at the University of Neuchâtel:[3] (Llamas Martínez 2018; Sánchez Jiménez 2017).

The ADSO project collected Golden Age Spanish sonnets from 52 different poets. These sonnets were automatically annotated with metrical information. Based on the analysis of these texts, lead researcher Navarro-Colorado drew conclusions about each author's preferred rhythmic patterns when writing sonnets (Navarro-Colorado 2015, 2016). As part of the studies conducted in the POSTDATA project, a larger corpus of Spanish sonnets was collected. This featured poems from a broader time period —from the 15th to the 19th centuries—and drew on the work of more than 1000 authors. Researchers were able to use this corpus to analyze metrical patterns in post-Golden Age sonnets, resulting in some interesting findings (Ruiz Fabo et al. 2020). Sánchez Jiménez (Sánchez Jiménez 2017) and Llamas (Llamas Martínez 2018) have also studied a corpus of Lope de Vega and Góngora romances to detect metrical differences between authors working in this subgenre. In short, quantitative metrical research in Spanish poetry has mostly focused on sonnets and romances and neglected the analysis of other subgenres. Similarly it has not considered the entire poetic works of any author.

3 Our Study

This study aims to contribute to the growing number of quantitative metrical studies about the metrical choices of four canonical early modern Spanish poets. To this end, we explore the complete metrical profile of each author

---

[1] ADSO stands for "Análisis distante del soneto castellano de los Siglos de Oro" ("Distant Reading Approach to Golden Age Spanish Sonnets"). Led by Borja Navarro Colorado, this project was funded by a Fundación BBVA Digital Humanities grant. For more information, please visit the project website: http://adso.gplsi.es/index.php/en/adso-project/ [accessed: 19/11/2020].

[2] POSTDATA stands for "Poetry Standardization and Linked Open Data". Led by Elena González-Blanco, this project received funding from the European Research Council. For more details, visit http://postdata.linhd.uned.es/ [accessed: 19/11/2020].

[3] This project was led by Antonio Sánchez Jiménez and funded by the Fonds National Suisse de la Recherche Scientifique. For more information, please see: http://www.snf.ch/fr/encouragement/projets/projets-toutes-les-disciplines/Pages/default.aspx [accessed: 19/11/2020].

without restricting our analysis to any specific subgenre. This approach is made possible by the development of Rantanplan (De la Rosa et al. 2020a), which performs metrical analysis not only on hendecasyllables but on every type of Spanish verse. Moreover, Rantanplan's new feature of stanza and structure identification enables us to analyze the entire range of strophic and structural usage for each author.

3.1 Dataset

As an essential first step in this project, we obtain the complete poetic works of several early modern Spanish poets. In this regard, one of the limitations of quantitative approaches to Spanish literature has been the scarcity of digitized texts in suitable form. Both digital editions and repositories are lacking. As a result, our selections are based on the relevance of each poet's works in the early modern Spanish period and their digital availability. Our study uses the complete poetic works of four such authors: Garcilaso de la Vega (1503–1536), Fernando de Herrera (1534–1597), Luis de Góngora (1561–1627), and Lope de Vega (1562–1635). Garcilaso and Herrera are considered Renaissance poets while Góngora and Lope are major representatives of the Baroque style. In the case of three of these authors, their entire poetic corpus is available online; the corpus of the fourth poet (Herrera) has not yet been published. For Garcilaso and Lope, we use the source texts provided by Wikisource[4] and Cervantes Virtual library respectively.[5] For Góngora's poetry, we use the *gongocorpus*, a TEI-xml format corpus provided by Averell,[6] a tool developed within the POSTDATA project (De la Rosa et al. 2020b).[7] For Herrera's corpus, we use the complete poetic works digitized within Hernández-Lorenzo (2020).[8]

   In order to prepare our texts for automatic metrical annotation —and to ensure there are no orthographic issues which could affect the analysis—, we standardize the orthography to contemporary Spanish and remove character names from dialogues along with rare symbols. We also exclude texts or fragments in languages other than Spanish from our analysis.

3.2 Methodology

To metrically analyze our corpora of early modern Spanish poets, we rely on digital humanities quantitative methodologies, as described in Section 2. We arrange for automatic scansion and syllabification with the open-source Python package Rantanplan (De la Rosa et al. 2020a). We also use Rantanplan's new

---

[4] Available at: https://es.wikisource.org/wiki/Autor:Garcilaso_de_la_Vega [accessed: 20/11/2020].

[5] Available at http://www.cervantesvirtual.com/obra-visor/poesias-liricas–0/html/ff775820-82b1-11df-acc7-002185ce6064.html [accessed: 19/11/2020].

[6] Available at https://github.com/linhd-postdata/averell [accessed: 19/11/2020].

[7] These source texts are in fact the digital editions of Góngora's poems prepared in the acclaimed POLEMOS project: http://obvil.sorbonne-universite.site/corpus/gongora/ [accesed: 11/11/2020].

[8] These texts were digitized using the most complete and widely respected edition of Herrera's poems (Herrera 1975). As some of his works raise questions about authorship and others contain variants of the same text, we limit our analysis to poems and versions of texts published in the author's lifetime in *Algunas obras* (Herrera 1582).

Author	Poems	Words
Garcilaso	60	26.182
Herrera	91	18.685
Gongora	320	74.564
Lope	51	21.707

Table 1: Corpora of authors in our study including the number of poems and words in their complete poetic works

stanza and structure identification feature, which works as follows: leveraging information about the metrical syllables and stressed endings in each line, our algorithm assigns a customizable window to each rhyme per line. Both metrical lengths and rhyme patterns are then fed to a pattern-matching algorithm which uses recurring expressions to search for matches within a growing knowledge base of structures and stanzas. The most comprehensive match is then returned in both machine and human-readable formats, enabling scholars to pursue further investigations.

We annotate each plain text corpus in two different formats: the first contains only stress information while the second combines stress and stanza annotation. We then use a Python script to extract and calculate quantitative metrical data for each author and compare the results, considering stress position, verse length, and stanza types.

4   Quantitative Metrical Analysis

A quantitative metrical analysis can give us a broad picture of a poet's metrical choices and help determine whether they arise from a particular poetic school or from individual preferences. In this paper, we first compare the distribution of line stresses across the complete poetic corpora of Garcilaso, Herrera, Góngora, and Lope. By analyzing recurring stress patterns and the proportion of particular stress positions in each corpus, we determine the authorship features of the four poets. We then turn to line lengths to assess how these differ for the individual authors; these characteristics can also help identify some authorship traits. Lastly, we analyze stanza use in the poetic works of the four authors. A quantitative analysis of the distribution of stanza types across the corpora can improve our understanding of these authors' use of different poetic forms.

4.1   Stress Positions

This section analyzes the distribution of stressed metrical positions across the four corpora. We refer here to metrical positions and not to syllables because certain phenomena may affect the one-to-one correspondence of these two elements. The use of a synaloepha, an extremely common feature in Spanish poetry, may, for example, cause two syllables to count as one metrical position.
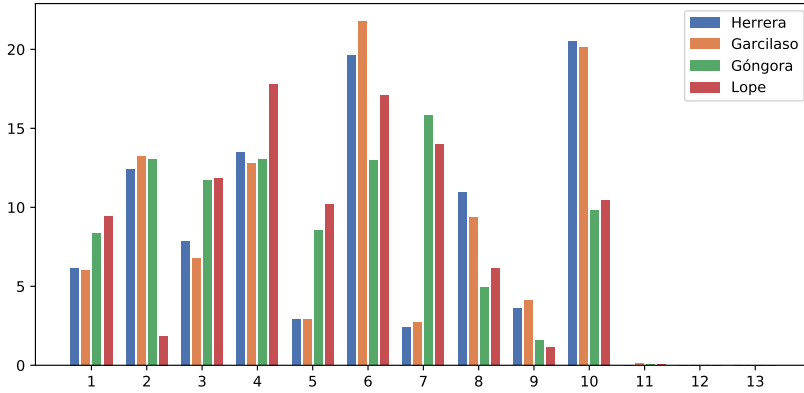
Figure 1: Distribution of stressed metrical positions per line in the complete poetic works of Garcilaso, Herrera, Góngora, and Lope

Rantanplan (De la Rosa et al. 2020a) detects these phenomena and calculates the number of metrical positions accordingly.

On the one hand, this comparison of the distribution of stress positions across the corpora highlights the poetic styles used by these authors and the literary schools to which they belong. On the other, the data reveal these poets' individual preferences regarding metrical position stress; these preferences are not strictly related to the style that they follow.

As shown in Figure 1, based on their stress patterns, the four authors can be split into two groups that respectively represent the Renaissance and the Baroque styles: Garcilaso and Herrera show similar patterns and are both Renaissance figures; Góngora and Lope constitute the Baroque group.

The most striking difference between the two groups relates to the frequency of stressed syllables in the seventh and tenth positions. Góngora and Lope often place a stress in the seventh position, while Garcilaso and Herrera tend to avoid this; the opposite scenario applies in the tenth position, which is strongly stressed in the corpora of Garcilaso and Herrera but is not similarly prominent in the works of the other two authors. Besides showing a marked tendency to stress the tenth position, the Renaissance group are also prone to stressing the sixth position. The co-occurrence of stresses in the sixth and tenth positions suggests that Garcilaso and Herrera mostly composed hendecasyllables *a maiore*. The popularity of hendecasyllables in these two authors' corpora is also evident when considering line-length alone, as we will see in Section 4.3. The predominance of the *a maiore* pattern corroborates the findings of previous traditional and computational studies of the Renaissance hendecasyllable (Domínguez Caparrós 2014; Henríquez Ureña 1919; Navarro-Colorado 2015, 2016; Ruiz Fabo et al. 2020). In fact, there is a clear preference for patterns based on the 6-10 scheme (e.g. 2-4-6-10, 2-6-10) among Renaissance poets (for a detailed account, see Navarro-Colorado (2016)). This preference decreased in Baroque era when the most frequent pattern was 2-4-8-10 (Navarro-Colorado 2016).
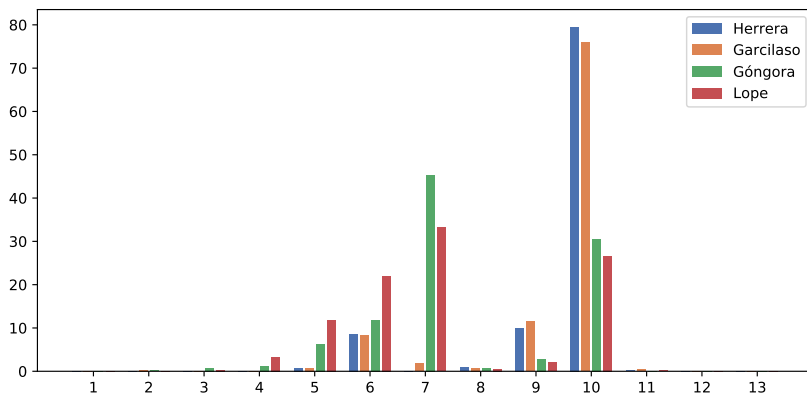
Figure 2: Distribution of line lengths based on the number of metrical positions in the poetry of Garcilaso, Herrera, Góngora, and Lope

By determining the stress pattern for individual authors, we can also identify distinctive features in their work. Lope's poetry stands out for its avoidance of a stress in the second metrical position in favor of one in the third position, while Góngora and the two Renaissance poets seem to prefer to stress the second position. This difference between Góngora and Lope can be explained by considering that, as Navarro-Colorado (2016) observe, the stress in the third position gradually became more widespread during the Baroque period and was a particularly prominent trait in the later stages of this era. Since Góngora is one of the first Baroque poets, this practice is not yet visible in his verse, while it is evident in a later poet like Lope.

A further trend that distinguishes Lope from Góngora is that the former prefers to stress the sixth position while the latter opts more often for a stress in the seventh position.

Finally, even though Garcilaso and Herrera turn out to have fairly similar metrical patterns, they each have some distinctive traits: for Garcilaso, the second, sixth, and ninth positions are stressed slightly more frequently than they are for Herrera; on the other hand, the third, fourth, and eighth positions are stressed more often for Herrera than they are for Garcilaso.

4.2 Line Lengths

Measuring the distribution of line lengths allows us to identify the most represented poetic forms in each of the corpora. Line lengths are calculated based on the number of metrical positions per line. As shown in Figure 2, we can identify two groups based on these line-length distributions: Renaissance and Baroque poets.

The clearest distinction between the two groups concerns the tenth metrical position. In the Renaissance group, the vast majority of lines have ten metrical positions, which suggests that these authors mostly wrote hendecasyllables. In contrast, Baroque authors show no preference for any line length; their

poetry contains lines of diverse lengths of which the most common has seven metrical positions. Lope's poetry shows the greatest variation in terms of line length.

4.3 Stanza Use

In terms of stanza detection, we obtain quantitative results by using Rantanplan (De la Rosa et al. 2020a) to calculate the proportion of each type of stanza for each author as well as the total number of stanzas in his work (see the repository table mentioned in the Appendix). Of the 47 known stanza types, 20 are not used by any the four early modern Spanish poets.[9] Some of these types including the *estrofa manriqueña, copla de arte mayor, copla de arte menor, copla real, copla mixta, décima antigua,* and *cuaderna vía* are typical of the medieval period. Another undetected type, the haiku, was not employed until recently in Spanish poetry.[10]

The rest of the stanza types are used by at least one author; only five, i.e. sonnets, couplets, *octavas, serventesios,* and *tercetos* are used by all four poets. Garcilaso and Herrera write *liras,* whereas Góngora and Lope favor *sextetos liras.* In Lope's case especially, this form comprises 17.05% of detected stanzas. The two Baroque poets also use some popular stanza forms such as *seguidillas, septillas* and *romances,* which are not detected among the Renaissance poets. The *terceto* is the most prevalent type for three of the authors, while more than 50% of the stanzas detected in Góngora's work are *cantares,* also known as *coplas.* Another stanza quite typical of the period, the *octava real,* is heavily used by Garcilaso and Góngora but not by Herrera and Lope. These results accord with our earlier observations about stress positions and line lengths. In particular, the popular stanza forms of the Baroque era and the *cantares* are non-hendecasyllabic. All this correlates with the reduced use of this type of verse by Góngora and Lope.

Of the most popular stanza types, sonnets and *tercetos* are of particular interest given their prevalence in the works of many of our authors. It is, thus, worth taking a closer look at the use of these two forms.

Figure 3 compares the relative frequency of sonnets across the works of the four poets. Lope favors this stanza type, which constitutes over 17% of his detected stanzas. Herrera is the next most frequent user with a rate of nearly 15%. In contrast, sonnets are far less common for Garcilaso and especially Góngora, with a proportion below 10% in both cases.

As noted above, the *terceto* is the most frequent stanza type in the work of three of the four poets. However, the extent of its use varies among those authors. While for both Garcilaso and Herrera, nearly 70% of detected stanzas

---

[9] These undetected types are the *copla de arte mayor, copla de arte menor, copla real, cuaderna vía, décima antigua, endecha real, estrofa manriqueña, estrofa sáfica, haiku, octava mixta, octavilla, ovillejo, quinteto, seguidilla compuesta, seguidilla gitana, septeto lira, sexteto, silva arromanzada, soleá,* and *terceto monorrimo.*

[10] Notably, other stanzas and structures undetected among the four poets include hendecasyllabic stanzas and structures such as the *estrofa sáfica, terceto monorrimo, quinteto, sexteto, septeto lira, endecha real,* and *silva arromanzada.* Regarding popular stanzas and structures, the *octavilla, ovillejo, seguidilla compuesta, seguidilla gitana,* and *soleá* are also unused.
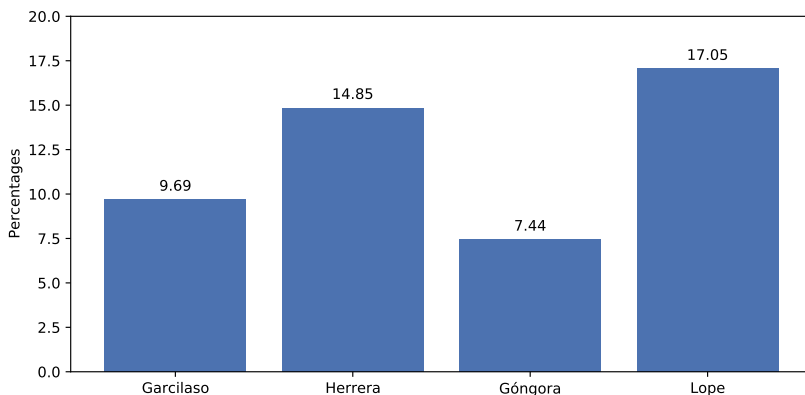
Figure 3: Proportion of sonnets used. Lope relies on sonnets most often (in more than 17% of his detected stanzas) followed by Herrera (almost 15%). Garcilaso and Góngora both employ this structure to a lesser extent (in less than 10% of detected stanzas in both cases).

are *tercetos*, over 78% of Herrera's poetic output is identified as such. On the other hand, this form is far less common for the two Baroque poets. Indeed, *tercetos* account for far less than 50% of Lope's works and do not even represent 3% of the stanzas detected in Góngora's works.

5   Conclusions and Future Work

This paper has applied a quantitative methodology to the metrical analysis of Spanish poetry, with a focus on four representative authors of the early modern period. We have relied specifically on the application of our automatic scansion and syllabification tool, Rantanplan. Thanks to the development of a new stanza identification feature in this software, we were able to perform a complete analysis of the metrical choices of the selected authors. By analyzing quantitative data about stress positions, line lengths, and stanzas, we aimed to model their metrical authorial profiles. Our results confirm previous hypotheses and provide new insights.

The predominance of *a maiore* patterns and the preference for patterns based on a 6-10 scheme correlate with the findings of earlier traditional and computational research about the Renaissance hendecasyllable. In addition, even where differences in stress positions delimit two distinct group of poets, i.e. Renaissance and Baroque figures, there are enough unique features to identify each of the authors.

Regarding line length, our results show that the two Renaissance authors mostly write in hendecasyllables while their Baroque counterparts exhibit a wider range of verse types. This is even more apparent in our quantitative stanza analysis, which confirms that shorter-verse stanzas become more popular among the Baroque poets;*cantares* are, for example, strikingly frequent
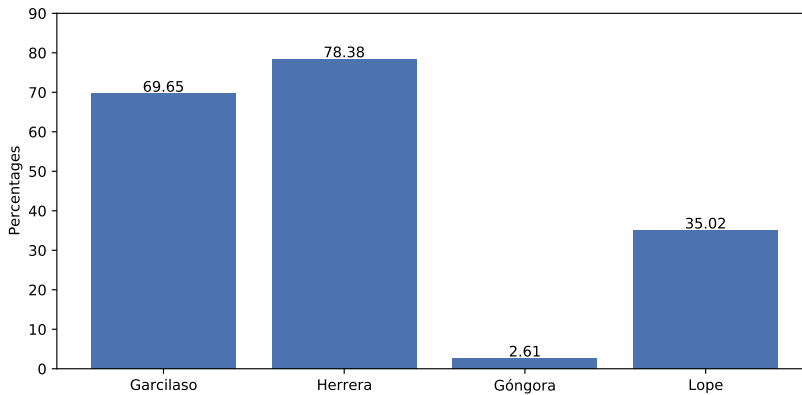
Figure 4: Proportion of *tercetos* used. Herrera most clearly favors this stanza form (in more than 78% of his detected stanzas), followed by Garcilaso (almost 70%). The proportion oftercetos then drops for Lope (to around 35%) and plunges even further for Góngora (to less than 3% of his detected stanzas).

in Góngora's works. Likewise, a close examination of the most popular stanza, the *tercetos*, makes clear that this generally hendecasyllabic stanza was far less common among Baroque than among Renaissance authors.

Significantly, all these findings are restricted to a small number of authors. As such, in future work, we would like to expand the corpora of poets with the goal of corroborating the data obtained in this study with those for other Renaissance and Baroque authors.

Acknowledgments

References

Agirrezabal et al (2013). "ZeuScansion: A Tool for Scansion of English Poetry". en. In: *Journal of Language Modelling* 4.1, p. 3. ISSN: 2299-8470, 2299-856X. DOI: 10.15398/jlm.v4i1.102. URL: http://jlm.ipipan.waw.pl/index.php/JLM/article/view/102 (visited on 07/03/2020).

Burrows, John Frederick (2004). "Textual analysis". In: *A Companion to Digital Humanities*. Ed. by S. Schreibman, R. Siemens, and J. Unsworth. Oxford: Blackwell, pp. 323–347.

De la Rosa et al. (2020a). "Fast and Accurate Syllabification and Scansion of Spanish Poetry". In: *Revista de Procesamiento del Lenguaje Natural* 65.

De la Rosa et al. (2020b). "PoetryLab as Infrastructure for the Analysis of Spanish Poetry". In: *Proceedings of CLARIN Annual Conference 2020*, pp. 82–87. URL: https://office.clarin.eu/v/CE-2020-1738-CLARIN2020_ConferenceProceedings.pdf.

Domínguez Caparrós, J. (2014). *Métrica española*. Madrid: UNED.

Gervás, Pablo (2000). "A Logic Programming Application for the Analysis of Spanish Verse". In: *Computational Logic – CL 2000*. Lecture Notes in Computer Science. Springer. DOI: 10.1007/3-540-44957-4_89.

Henríquez Ureña, P. (1919). "El endecasílabo castellano". In: *Revista de Filología Española* 6, pp. 132–157.

Hernández-Lorenzo, Laura (2020). "Los textos poéticos de Fernando de Herrera. Aproximaciones desde la estilística de Corpus y la Estilometría". PhD thesis. Spain: Universidad de Sevilla.

Herrera, F. de (1582). *Algunas obras de Fernando de Herrera*. Casa de Andrea Pescioni.

Herrera, F. de (1975). *Obra poética*. Ed. by J. M. Blecua. Boletín de la Real Academia Española.

Jockers, Matthew L. (2013). *Macroanalysis. Digital Methods and Literary History*. University of Illinois Press.

Llamas Martínez, Jacobo (2018). "Métrica, ritmo acentual y autoría en la poesía española del Siglo de Oro". In: *Monográfico de Arte Nuevo* 5, pp. 49–214.

Moretti, Franco (2003). "Graphs, Maps, Trees: Abstract Models for Literary History 1". In: *New Left Review* 24, pp. 67–93.

Moretti, Franco (2013). *Distant Reading*. London: Verso.

Navarro-Colorado, Borja (2015). "A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects". In: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. Denver, Colorado, USA: Association for Computational Linguistics, pp. 105–113. DOI: 10.3115/v1/W15-0712. URL: https://www.aclweb.org/anthology/W15-0712 (visited on 05/17/2020).

Navarro-Colorado, Borja (2016). "Hacia un análisis distante del endecasílabo áureo: patrones métricos, frecuencias y evolución histórica". In: *Rhythmica*.

Navarro-Colorado, Borja (2017). "A metrical scansion system for fixed-metre Spanish poetry". In: *Digital Scholarship in the Humanities* 33.1, pp. 112–127.

Plecháč et al., ed. (2019). *Quantitative approaches to versification*. Institute of Czech Literature of the Czech Academy of Sciences.

Ruiz Fabo et al. (2020). "The Diachronic Spanish Sonnet Corpus: TEI and linked open data encoding, data distribution, and metrical findings". In: *Digital Scholarship in the Humanities*. DOI: https://doi.org/10.1093/llc/fqaa035.

Sánchez Jiménez, Antonio (2017). "Acentos contiguos en los romances de la Arcadia (1598), de Lope de Vega". In: *Atalanta: revista de las letras barrocas* 5.1, pp. 5–61.

Schreibman, S., R. Siemens, and J. Unsworth (2004). *A Companion to Digital Humanities*. Oxford: Blackwell.

Appendix

To see a complete table of relative values, i.e. percentages, for the results of stanza and structure detection across the four corpora, please visit the repository at https://github.com/linhd-postdata/Plotting_Poetry2020. This repository also contains the corpora used in this project, with the exception of Herrera's texts, which have not yet been published. Herrera's corpus annotations are, however, included.