Stylistic Change in Early Modern Spanish Poetry Through Network Analysis (with an Especial Focus on Fernando de Herrera’s Role)

Laura Hernández-Lorenzo

Accepted: 15 November 2021 / Published online: 1 April 2022 © The Author(s) 2022

Abstract

Early Modern Spanish literature, also known as Golden Age Spanish literature, is a well-established period in the History of Spanish literary tradition, covering from the beginning of the sixteenth century to the end of the seventeenth century, from Renaissance to Baroque works. In the case of poetry, the stylistic change from one aesthetic to the other has been examined from a historical point of view, with some scholars considering a transitional group of poets between both stylistic movements, and many pointing to the writer Fernando de Herrera (1534–1597) as a bridge between the Renaissance style of Garcilaso de la Vega (1501–1536) and the Baroque of Luis de Góngora (1561–1627). This paper examines the stylistic change from Renaissance to Baroque in Spanish poetry and Herrera’s place on it from a quantitative, computational point of view, applying a methodology which combines Stylometry and Network Analysis to a big corpus of poems and authors writing in this period. The resulting stylometric network for Renaissance and Baroque poetry proves there is a computationally measurable chronological evolution in the texts of this period, as well as a change of style from Renaissance to Baroque, and supports the existence of a transitional group of poets between these two literary movements, one of which is Fernando de Herrera.

Keywords Spanish poetry · Renaissance · Baroque · Fernando de Herrera · Stylometry · Network analysis

Laura Hernández-Lorenzo
laurahlr@gmail.com

1 Institute of Polish Language PAS, Krakow, Poland
Introduction

Early Modern Spanish literature, also known as Golden Age Spanish literature (‘Siglo de Oro’), is a well-established period in the History of Spanish literary tradition. It is well known it covers literary works from the beginning of the sixteenth century to the end of the seventeenth century, from Renaissance to Baroque works. In the case of poetry, the stylistic change from one aesthetic to the other has been examined from a historical point of view, with some scholars considering a transitional group of poets between both stylistic movements (López Bueno, 2000, 2006), and others referring to this transition as Mannerism (Orozco 1971, 1981; Lara Garrido, 1979, 1980).

According to some literary critics, one of the key poets in this stylistic evolution is Fernando de Herrera (1534–1597), who has been considered to act as a bridge between the Renaissance style of Garcilaso de la Vega (1501–1536) and the Baroque one of Luis de Góngora (1561–1627). The posthumous edition of his poetry, Versos de Fernando de Herrera (1619), is considered more Baroque than the one published during his life, and a precedent of culteranismo. Apart from Oreste Macrì’s notes on this topic (1972), there are no studies focusing on it in a systematic way. Likewise, research on the stylistic evolution in Early Modern Spanish poetry –especially on the evolution from Renaissance to Baroque– will doubtlessly benefit from a quantitative approach which allows the researcher to work with a big corpus of texts and authors.

The present study addresses the stylistic change from Renaissance to Baroque in Spanish poetry –as well as the role of Herrera in it– taking advantage of the latest developments in Digital Humanities quantitative approaches to literature research, and especially of the combination of Stylometry with Network Analysis. In the first two sections of this paper, a double state-of-the-art is presented. This includes an account of the principal studies examining the stylistic change from Renaissance to Baroque and the discussion on Fernando de Herrera’s role, combined with an overview of quantitative approaches in Digital Humanities, with an especial focus on Stylometry and how computer analysis may contribute to understand literary history. The next two sections introduce the dataset and the methodology used in this study, followed by the results and discussion, and finally conclusions and future work. An appendix at the end of the paper includes complementary materials.

Early Modern Spanish Poetry and the Discussion on Fernando de Herrera’s Role

The stylistic change from Renaissance to Baroque in Spanish poetry has been examined from a historical point of view in relevant studies. Of particular importance among these are the volumes written or directed by Begoña López Bueno (2000, 2006). Both La poética cultista de Herrera a Góngora and La renovación poética

1 Traditionally, the term culteranismo has been used in Spanish literature scholarship to refer to the new poetic style created by Luis de Góngora, who is, according to experts, “el autor barroco por excelencia” (Blanco 2016: p. 16).
del Renacimiento al Barroco delve in the different stages and authors in this period, while they argue for the existence of a transitional group of poets between the Renaissance and the Baroque stylistic movements. The latter idea was previously supported by Emilio Orozco (1971, 1981) and José Lara Garrido (1979, 1980) in their studies about Mannerism.

According to Spanish literature scholars, an important date for the transition between the Renaissance and the Baroque would be 1580, with the publication of Anotaciones a la poesía de Garcilaso de la Vega by Fernando de Herrera, and the beginnings of Lope de Vega’s and Luis de Góngora’s writing. Another turning point was the publication in 1605 of the Flores de poetas ilustres, a poetic anthology prepared by Pedro de Espinosa. The Flores reflects the aesthetic changes in the poetic genre, as well as the evolution and renovation towards culteranismo and a Baroque poetry (Ruiz Pérez, 2000). In this context, Molina Huete remarks that “las Flores de 1605 […] contribuyen de esta manera a perfeccionar el dibujo de la orografía lírica de un periodo impreciso que, sin entrar en complicadas consideraciones críticas, sencillamente denominamos manierismo” (Molina Huete, 2003, p. 98).

The Sevillian poet Fernando de Herrera, also known as the ‘Divine’, has been pointed out as a crucial author in the transition from the Renaissance to the Baroque. Some recognised scholars consider him to act as a bridge between the Renaissance style, embodied by Garcilaso de la Vega, and the Baroque style, whose ultimate referent would be Luis de Góngora. Accordingly, Vilanova regards Herrera as “el verdadero precursor del culto a la erudición poética, del hermetismo y de la oscuridad barroca de don Luis de Góngora” (1951: p. 710), and he adds: “su doble faceta de poeta intimista y cantor sacro y heroico, acusan un progreso técnico de tal envergadura y una anticipación tan manifiesta del espíritu romántico del barroco, que le convierten en un hito crucial de nuestra historia literaria” (1951: p. 714). Similarly, another scholar, Valbuena Prat, stresses Herrera’s role as “un punto medio entre el estilo de Garcilaso y el de Góngora […] el puente necesario entre la escuela de Garcilaso y el culteranismo” (Valbuena Prat, 1960, I: pp. 545–552). However, María Teresa Ruestes and Begoña López Bueno emphasize the complexity of The Divine’s part in the stylistic change. According to Ruestes, Herrera has “un lugar eminentemente y una excepcional significación en la poesía de signo cultista de la lírica española del Siglo de Oro” (Ruestes, 1986: p. XXXVI). And López Bueno claims that the progression Garcilaso-Herrera-Góngora especially applies to a formal level, that is, through an intensification of rhetorical devices, a growing morphosyntactic complexity and the usage of erudite lexical forms (López Bueno, 2000: p. 33).

Most importantly, this vision of Herrera as a transitional or middle point between the two aesthetics movements is intrinsically connected to the authorial

---

2 For a complete study on the controversy generated by Herrera’s Anotaciones see Montero’s monograph (Montero 1987).
3 Góngora especially will become the most notorious poet of culteranismo and, as a result, a big exponent of the Baroque style and the new poetry (Blanco 2016).
4 In this same year the second part of the Romancero general and the first part of the Quixote were published. Considering that the Arte nuevo de hacer comedias was written in the following years, 1605 emerges as a key date for the aesthetic change across different literary genres (Ruiz Pérez 2000).
problem surrounding his poetic works, known as the ‘textual drama’. This controversy revolves around the differences between the only edition of Herrera’s poetry published during his life –titled *Algunas obras* (1582) and known as H– and the posthumous edition –titled *Versos de Fernando de Herrera* (1619) and known as P–. Whereas *Algunas obras* includes a selection of The Divine’s poems prepared by him, *Versos* was published some years after Herrera’s death by the painter Francisco Pacheco (1564–1644). The posthumous edition includes new poems and different versions of the ones published in 1582. The existence of significant differences between the two editions has led some recognized scholars to reject Herrera’s full authorship of the new poems and variants of old ones published in the posthumous edition (Blecua, 1958; Cuevas, 1985; Kossoff, 1965). In spite of this, other erudite academics defend Herrera’s full authorship and justify these differences as the result of a progression towards a more Baroque style (Battaglia, 1954; Macrí, 1972; Pepe Sarno, 1982).

This complex authorship problem has already been examined through styliometric and non-traditional authorship attribution methods (Hernández Lorenzo, 2019), with results supporting the authenticity of the posthumous edition. Nonetheless, what is of interest for the present study is how the two different positions on the authorial controversy are connected to different views on Herrera’s role in Spanish poetry. In this sense, those scholars supporting Herrera’s authorship defend the existence of an evolution of his style towards Baroque; in contrast, the ones rejecting *Versos*’ authenticity see Herrera as a fairly Renaissance poet. Even more, scholars from both sides of the debate seem to agree on the more Baroque component of the posthumous edition’s poems. According to literary scholars, this would be the biggest difference between the poems published in 1582 and those published in 1619. Accordingly, Salvatore Battaglia refers to how *Versos* reflects the stylistic changes Herrera introduces towards the Baroque as “la funzione riformatrice che ha esercitato con la sua sensibilità lessicale e sintattica il poeta sivigliano nell’intero circolo della lingua spagnola” (Battaglia, 1954: p. 87). José Manuel Blecua explicitly mentions the Baroque component of the posthumous poems: “el barroquismo que caracteriza a los textos de Pacheco” (Blecua, 1958: p. 391). And Cristóbal Cuevas stresses the stylistic similarities of *Versos*’ poems with typical Baroque stylistic resources, and more specifically, Góngora’s ones: “muchos poemas de 1619 sapiunt Gongoram” (Cuevas, 1985: p. 99).

To this date, we do not count with studies addressing this issue in a systematic way, with the exception of Macrí’s notes (1972). Likewise, research on the stylistic evolution in Early Modern Spanish poetry –especially on the evolution from Renaissance to Baroque– will doubtlessly benefit from a quantitative approach which allows the researcher to work with a big corpus of texts and authors.

---

5 Pacheco is especially famous in the History of Spanish Art for being the teacher and father-in-law of the great Spanish painter Diego Velázquez (1599–1660).
Quantitative Methods for (Spanish) Literature Research

The growing field of what we call today Digital Humanities (Schreibman et al., 2004) has led to new approaches in Literature Studies. These imply new methodologies, perspectives, and tools for the study of literature, which can help literature scholars in their research, as well as shed new light on when traditional methodologies such as close reading or Philology are not able to do so, and even change our perception of literary works and Literary History. One of the approaches that this digital turn has boosted is the application of quantitative methods for literature research.

According to Hoover (2008), modern quantitative studies of literature began about 1850, but recent advances –such as the significant growth in the availability of electronic texts, increasingly sophisticated statistical techniques, and the emergence of more powerful computers– have led to more accurate analysis. As Hoover points out, “Quantitative approaches to literature represent elements or characteristics of literary texts numerically, applying the powerful, accurate, and widely accepted methods of mathematics to measurement, classification, and analysis” (Hoover, 2008). Although they are most naturally associated with questions of authorship and style, “they can also be used to investigate larger interpretative issues like plot, theme, genre, period, tone, and modality” (Hoover, 2008).

Quantitative approaches have become especially prominent after the publication of “Graphs, maps and trees” by Franco Moretti (2005). In the same years, Matthew Jockers published Macroanalysis: Digital Methods and Literary History (Jockers, 2013). He claims that “what we have today in terms of literary and textual material and computational power represents a moment in revolution in the way we study the literary record […] large scale text analysis, text mining, ‘macroanalysis’, offers an important and necessary way of contextualizing our study of individual works of literature” (Jockers, 2013, p. 171). These approaches have been reinforced by the foundation of the Stanford Literary Lab,7 where researchers study literature using this methodology, producing results that are published in the Lab’s website as Pamphlets.

One of the disciplines that tend to be included among quantitative approaches in Digital Humanities is Stylometry.8 The term appeared for the first time in Lutosławski’s “Principes de Stylométrie…” (1898), and, from an etymological point of view, is formed by two other words, style and (statistical) metrics. Stylometric methods are rooted in the assumption that it is possible to explore style and its different traits by measuring linguistic aspects of a text –usually the most frequent words– through the application of statistical measures. Historically, most of the

---

6 According to Moretti (2005), distant reading emerges as an opposite to close reading and as a way to bring more objectivity in Literature Studies. Instead of analysing one text or fragment in detail, he defends the use of a more abstract methodology which enables the researcher to work with huge amount of texts in order to see patterns and relations between them.

7 See website at: https://litlab.stanford.edu/ (accessed: 29/11/2020).

8 In A Companion to Digital Literary Studies, Hoover includes non-traditional authorship attribution and Stylometry in his chapter on “Quantitative Analysis and Literature studies” (Hoover 2008).
stylometric studies have dealt with authorship attribution problems. Since Mosteller and Wallace’s success with *The Federalist Papers* (1963), most stylometric studies rely on function words—that is, pronouns, articles, prepositions and conjunctions, among others. In order to quantify the style of a text, Stylometry generally works with high frequency items, which, in the case of words, are function words in a large proportion. In addition, in recent evaluation research most frequent words have turned out to be the most reliable feature (Hettinger et al., 2016).

Once the features have been selected, Stylometry research uses statistical and computational techniques to analyse this data. This generally involves the application of distance metrics, which measure and quantify the stylistic similarity or difference between texts (Juola, 2006; Rotari et al., 2020). Among these distances, the most popular are Delta (Burrows, 2002) and its adaptations. The resulting distance values are then visualised through a variety of methods, mainly unsupervised ones, such as cluster analysis or bootstrap consensus tree (Eder, 2013). The first one, hierarchical cluster analysis, “is a technique which tries to find the most similar samples (e.g. literary texts), and builds a hierarchy of clusters using a bottom-up approach” (Eder, 2017b: p. 51). In the second one, “In a very large number of iterations, the variables needed to construct a dendogram were chosen randomly, and a virtual dendogram for each iteration was generated. Next, these numerous virtual dendograms were combined into a single compact consensus tree” (Eder, 2013). Hence, the difference between both procedures is that the second one shows only the most stable relations between texts, that is, those which appear along several iterations. These methods have already being applied to authorship problems of some Spanish Early Modern works, such as the *Lazarillo* (de la Rosa & Suárez, 2016; Rißler-Pipka, 2016a), the Avellaneda’s *Quixote* (Blasco, 2016; Rißler-Pipka, 2016b), the plays *La conquista de Jerusalén* (Calvo Tello & Cerezo Soler, 2018), *Siempre ayuda la verdad* (García-Reidy, 2019), and the disputed plays by Moreto (Ulla Lorenzen et al., 2020) among others. There has been also some research on other stylistic phenomena—mainly literary genre— but to a lesser extent, in which the studies by the CLiGS project stand out (Calvo Tello, 2019), as well as those applied to the mythical fable (Rojas Castro, 2017), to the distinctiveness of Góngora and Picasso’s style (Rißler-Pipka, 2019) and to Middle Ages works (Fradejas Rueda, 2019).

In 2017, Maciej Eder published a new method for visualizing and analysing the stylistic and stylometric relations between texts through Network Analysis (Eder,

---

9 Against content words, function words form a close-class—difficult to expand with new additions—, are heavily grammaticalized and do not carry a lot of meaning in isolation (Kestemont 2014). These words present a number of advantages for Stylometry and non-traditional authorship attribution research: firstly, they are topic independent, since they are not strongly affected by a text’s topic or genre; secondly, they are much more frequent than content words, which makes them interesting from a quantitative and statistical point of view; finally, as their use seems to be less under the conscious control of the author, they are less prone to conscious manipulation and more resistant to imitation and plagiarism (Kestemont 2014).

10 According to John Burrows, the advantage of this method is that “it offers a rather harsh test of the questions to be considered and also […] the ‘family trees’ in which the results are displayed speak plainly for themselves” (Burrows, 2004).

11 This is defined by a consensus value, which is set to 0.5 by default in stylo (Eder et al., 2016).
Stylistic Change in Early Modern Spanish Poetry

While a consensus tree extracts only the strongest patterns and filters out weaker text similarities in order to detect the authorial signal, in Network Analysis these weaker connections are retrieved with the aim of addressing a variety of stylistic components beyond authorship attribution. For this reason, this method is particularly interesting for studying possible similarities between texts in terms of the influence of the author’s gender, the distinction between literary genres or other factors such as chronology and literary history.

The present paper aims to explore how computer analysis—and specially Stylometry and Network Analysis—can help to understand literary history and the evolution of poetic style in Early Modern Spanish poetry. An additional goal is to shed new light on the role played by Fernando de Herrera in this period: can he be considered as a transitional poet between Spanish Renaissance and Baroque? Are the poems in his posthumous edition more Baroque than the rest of his works?

Dataset

The first necessary step to carry out this project was to prepare a representative corpus of Early Modern Spanish poetry which included Herrera’s undoubted poems as well as the poems from the posthumous edition. For Spanish literature, an important limitation to carry out a quantitative study is the scarcity of digitalized texts in a suitable format to be used; that is, TEI-XML, plain text or even PDF. Despite the efforts of some institutions as the Spanish National Library or the Cervantes Virtual, a big proportion of the digitalized documents are scanned images prepared for human reading, not for the application of Natural Language Processing technologies. And the use of OCR (Optical Character Recognition) returns such noisy results that it is often preferable to transcribe the text from scratch. In short, Spanish literature suffers from few resources and repositories—especially when compared to English or German—.

In this work, we have used the corpus of Golden Age Spanish sonnets prepared in the ADSO project (Navarro-Colorado et al., 2016). This includes 52 Early Modern Spanish poets of the sixteenth and seventeenth centuries, from Juan Boscán (1490–1542) and Garcilaso de la Vega to the latest Baroque of Sor Juana Inés de la Cruz (1648–1695). Following the size constraints to detect authorial signals properly (Eder, 2017a), not every ADSO poet will be used, but only those whose collection of sonnets reach at least 1800 words. Íñigo López de Mendoza, Marquis of Santillana (1398–1458) was excluded for two reasons: firstly, Literature Criticism considers him a medieval or—at most—pre-Renaissance poet; secondly, his poems

12 As Eder points out, “The difficulties with separating one specific signal suggest that a text (written or spoken) is a multilayer phenomenon, in which particular layers are correlated. These layers include authorship, chronology, personality, gender, topic, education, literary quality, translation (if applicable), intertextuality, literary tradition (e.g. sources of inspiration), and probably many more […]. In authorship attribution, this complex system of uncontrollable layers is a problem of unwanted noise, and in literary-oriented computational stylistics, an opportunity to see more” (Eder 2017b: p. 53).
are written in an old Spanish which presents important orthographic and linguistic differences compared to the language used by the rest of the authors.

For Herrera, instead of using the poems offered by the ADSO corpus, which are a mix of *Algunas obras* and *Versos* edition, we used all the sonnets in his undoubted work, that is, the ones published in 1582 and some dispersed poems, as prepared and digitalised in Hernández-Lorenzo’s dissertation,¹³ which uses the most reputed edition of Herrera’s poems (1975). With the aim of determining the role of the posthumous poems in the Early Modern period, *Versos* unique sonnets –the ones that have not been published before 1619 in any version– were extracted from the same source and collected in a separate file, titled *P2*.¹⁴ Additionally, the digitalization of Francisco Pacheco’s poems in Hernández-Lorenzo’s dissertation was used to obtain his sonnets, which were added to the dataset.

In order to make sure that the same norms of modernization have been used in the two corpora, they were contrasted using ‘oppose’ function in ‘stylo’¹⁵ (Eder et al., 2016). The resulting corpus is in Unicode UTF-8 plain text and contains 40 collections of texts from 40 authors plus the attributed work to Herrera (see “Appendix” for further details). The range of authors goes from poets living and writing at the beginning of the sixteenth century to authors writing at the end of the seventeenth century. The corpus is well-balanced in terms of genre, since all the texts belong to the same poetic genre, the sonnet.

**Methodology**

The methods used here belong to Digital Humanities quantitative approaches, with a focus on the joint possibilities of Stylometry and Network Analysis presented by Eder (2017b). Eder’s method is based on the implementation within stylo (Eder et al., 2016) of an automatic generated table of textual connections for every cluster analysis or consensus tree produced.¹⁶ This table and its information are compatible with the widely used Network Analysis software Gephi (Bastian et al., 2009). By importing the stylo table into Gephi using the “Edge table” option, a textual network is produced. In the resulting network, texts are represented as nodes and the relations between them as links between the nodes. The network is based on the stylo-metric analysis, but contains more connections between the texts that the original consensus tree.

¹³ See Hernández-Lorenzo’s unpublished dissertation at: https://idus.us.es/handle/11441/93465 (accessed: 26/12/2020).
¹⁴ With the aim of avoiding possible biases created by poems that appear in both editions, the analyses are restricted to the new poems of *P*.
¹⁵ Spelling and orthographic differences are one of the features that emerge when comparing corpora. Using oppose function in *stylo*, one of the preferred words for Herrera and Pacheco’s texts was *solo* (‘only’), whereas one of the avoided words –and preferred by ADSO corpus– was *sólo*. When the orthography of this word was standardized to *solo*, both *solo* and *sólo* disappeared from the preferred and avoided wordlists.
¹⁶ This table is a CSV which name ends with “_EDGES”.
In this study, a consensus tree of the Early Modern Spanish corpus was generated in stylo applying Cosine Delta on the interval from 100 to 1000 most frequent words in the corpus. Since the goal here is not to focus on the authorial signal, but in other type of stylistic connections between the texts and authors in the corpus, the selected interval has the advantage that the analysis will be mostly restricted to function words while including a significant amount of iterations and, thus, ensuring that the textual connections retrieved are particularly stable. Cosine Delta (Smith & Aldridge, 2011) was chosen as it is the best performing distance metric in Stylometric studies up-to-date (Evert et al., 2017; Ochab et al., 2019). As final step for the creation of the network, one of the force-directed layout was chosen, namely the algorithm Force Atlas 2 (Jacomy et al., 2014), the recommended one in Eder’s article (2017b).

Results and Discussion

Figure 1 shows the resulting network for Renaissance and Baroque Spanish poetry. Firstly, the degree partition was applied to the network to check whether it is balanced in terms of the number of connections each author has with others in the network. As observed, although some authors have a greater number of connections than others, in general, the network obtained shows a certain balance in this sense, which makes it easier to see relations between the texts.

Also, if we look at the list of authors, it is striking that the poets from the beginning of the sixteenth century are located in the lower left corner of the graph –this is the case of Juan de Boscán or Garcilaso de la Vega–, while the poets from the end of the seventeenth century –such as Sor Juana or Polo de Medina– are located in the upper right corner. This seems to indicate the influence of a chronological signal on the constitution of the network.

Exploration of a Possible Chronological Signal

With the aim of verifying the relevance that chronology may have in the network, the years of birth of the different authors were introduced. Figure 2 shows the network coloured according to this data. As a result, poets with earlier dates of birth appear in lighter colours, while those with later dates of birth receive darker colours. As can be seen, the obtained figure supports this hypothesis, since lighter colours –and therefore earlier dates– are observed in the lower left corner of the graph, which gradually vary to the darker colours in the upper right corner of the graph.

In addition to the graphical representation of dates of birth in Fig. 2, the chronological evolution in the network was evaluated applying statistical tests to the

17 Force-directed layout are particularly helpful in Network Analysis, because they “perform gravity-like simulations and pull the most connected nodes (i.e. the ones that have several links and / or their links are very strong) to the center of the network, while the least connected nodes are pushed aside” (Eder 2017b: p. 59).
following measures of centrality used in Network theory: \textit{closeness centrality}\textsuperscript{18} and \textit{harmonic closeness centrality}.\textsuperscript{19} The values of these centrality measures for the network elements were calculated by Gephi and collected in a data table. This table was exported to R, where a script ("Script Networks.R") was created to perform the statistical calculations (see “Appendix”).

The first step was to generate a histogram\textsuperscript{20} using the poets’ dates of birth. As observed in Fig. 3, they follow a normal distribution,\textsuperscript{21} that is, there is a smaller number of poets in both ends of the graph (with a very early or very late date of birth in relation to the total), and a larger number of poets in the centre –close to the mean–.

Secondly, with the aim of exploring whether there is a correlation between dates of birth and centrality in our network, the difference between the year of birth of each author and the mean was calculated and converted to an absolute value. Then, a linear regression\textsuperscript{22} was generated using these values in addition to those obtained with each of the network centrality measures used. The \textit{p}-value\textsuperscript{23} obtained with closeness centrality is 0.04431. In contrast, a \textit{p}-value of 0.1452 is obtained with harmonic closeness centrality. As a convention, \textit{p}-values are significant if they are below 0.05. Hence, the \textit{p}-value obtained for closeness centrality is fairly good, but not the one for harmonic closeness centrality. However, we have identified that one author, José de Litala y Castelví, was acting as an outlier. If this author is isolated and we repeat the analysis without him, \textit{p}-value for closeness centrality is 0.0004003, and for harmonic closeness centrality we get 0.001972. This supports the existence of a correlation between dates of birth and centrality in the network, reinforcing our hypothesis about the importance of chronology in the formation of the network. Since no chronological data of the authors was available for the computer when the network was created, it is proved that there is a time signal in the texts of this period which can be detected through computational analysis.

\textsuperscript{18} “[…] closeness centrality […] measures the mean distance from a vertex to other vertices” (Newman, 2010, p. 181).

\textsuperscript{19} This is a variant of the previous one (Marchiori & Latora, 2000; Rochat, 2009).

\textsuperscript{20} A histogram is a graph similar to a bar chart, but instead of representing two or more variables, it represents an estimate of the values distribution of only one variable.

\textsuperscript{21} A normal distribution is defined as follows:

\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

where \( \mu \) represents the mean of the distribution, \( \sigma \) represents the standard deviation, \( \sigma^2 \) represents the variance.

\textsuperscript{22} Given two-dimensional data belonging to two variables, the generation of a linear regression consists of looking for the straight line that best fits the original data.

\textsuperscript{23} The \textit{p}-value is the probability that an observed event is a statistical fluctuation. If a deviation from the expected occurs when handling a random variable, the \textit{p}-value is calculated to know whether this is a statistical fluctuation or a significant value. As a convention, it is stated that it would not be a statistical fluctuation if the \textit{p}-value obtained is less than 0.05. The lower the \textit{p}-value, the less likely it is to be a statistical fluctuation and the better the result.
Poetic Evolution and Modularity

Apart from investigating a chronological signal, the aim of the present study is to examine the extent to which the stylistic evolution in Early Modern Spanish poetry—and from the Renaissance to the Baroque—is reflected in the obtained network. As mentioned above, this topic was studied from a historical non-computational point of view by Begoña López Bueno in the aforementioned volumes, where she argues for the existence of a transitional poetic group between Renaissance and Baroque poetry, a theory previously defended by Orozco (1971, 1981) and Lara Garrido (1979, 1980) in their studies about Mannerism.

To check this, a community detection technique was applied to the network. In these methods, the network is divided into several groups or communities.
Fig. 2  Network analysis of Golden Age Spanish Sonnets using Cosine Delta. Ranking by writer date of birth. Light colours represent earlier dates of birth, whereas darker colours represent later ones. (Color figure online)

Fig. 3  Histogram featuring the authors in the network by date of birth
using criteria not specified by the researcher, but by the very constitution of the network:

[…] the number and size of the groups into which the network is divided are not specified by the experimenter. Instead they are determined by the network itself: the goal of community detection is to find the natural fault lines along which a network separates […]. search for the naturally occurring groups in a network regardless of their number or size, which is used primarily as a tool for discovering and understanding the large-scale structure of networks (Newman, 2010, pp. 357, 371).

Among the different community detection methods, the most used is modularity calculation. On top of that, a recent study has shown it to be the more efficient one (Ochab et al., 2019). This measure delimitates the different communities that form the network by calculating numerical values:

[…] modularity […] has a high value when many more edges in a network fall between vertices of the same type than one would expect by chance. […]. Thus one way to detect communities in networks is look for the divisions that have the highest modularity scores and in fact this is the most commonly used method for community detection (Newman, 2010, pp. 372-373).

In order to detect the most consistent communities for Early Modern Spanish poetry, modularity values for the obtained poetry network were calculated. With a modularity value of 0.569, three communities emerge (see Fig. 4). In the first community, in brown colour, Juan Boscán, Garcílaso de la Vega, Diego Hurtado de Mendoza, Gutiérrez de Cetina, Juan de Almeida, Francisco de FIGueroa, Juan de Timoneda, Hernando de Acuña and Pedro de Padilla are included. The second community is represented in light blue colour and is made up of Francisco de Aldana, Diego Ximénez Ayllón, Cervantes, Pacheco, Herrera, Juan de Jáuregui, Juan de Argüijo, Francisco de la Torre, Luis Carrillo y Sotomayor, Soto de Rojas, Luis Martín de la Plaza and Pedro de Espinosa. Finally, the third community, coloured in green, contains Francisco de Medrano, los hermanos Argensola, Lope de Vega, Tirso de Molina, Mira de Amescua, Francisco de Borja, Quevedo, Góngora, Trillo y Figueroa, Bernardino de Rebollo, José de Litala y Castelví, Agustín de Salazar, Luis de Ulloa y Pereira, Juan de Tassis y Peralta, Bocángel y Unzueta, López de Zárate, Sor Juana Inés de la Cruz and Polo de Medina.

At first glance, it draws our attention how the obtained communities seem to correspond to different chronological groups of authors, measured by dates of birth (compared Figs. 2 and 4). To verify this, an evaluation was performed in R (see script at “Appendix”). In this evaluation, boxplots showing modularity communities by authors’ dates of birth were generated (see Fig. 5).

Figure 5 confirms that the obtained modularity communities correspond to different chronological groups.

Notwithstanding the proven relevance of chronology, the question remains as to what extent the stylistic evolution from the Renaissance to the Baroque may be a significant factor in the Early Modern Spanish poetry network. The brown community
contains poets which are usually classified as Renaissance ones by Literary Criticism. This is the case of Boscán, Garcilaso, Hurtado de Mendoza, Gutierre de Cetina or Hernando de Acuña. In contrast, the green community includes poets traditionally seen as Baroque, such as Francisco de Quevedo, Góngora, Lope de Vega, Tirso de Molina, Luis de Ulloa y Pereira, Bernardino de Rebolledo, Bocángel y Unzueta, Polo de Medina and Sor Juana Inés de la Cruz. The third community, coloured in light blue and placed between the previous two, combines authors normally assigned to the Renaissance –Francisco de Aldana, Francisco de la Torre or Fernando de Herrera, among others– with poets typically considered as Baroque –e.g. Juan de Arguijo, Juan de Jáuregui or Miguel de Cervantes–. This last community can be related to the existence of a transitional poetic group such as the one defended by López Bueno in her aforementioned studies (López Bueno, 2000, 2006).

Interestingly enough, both Herrera and the unique poems of Versos appear in this intermediate community. This boosts the transitional role between Renaissance poetry –represented by Garcilaso– and the Baroque one –with Góngora as its highest expression– assigned to Herrera and defended, as aforementioned, by some literature scholars, such as Dámaso Alonso, Antonio Vilanova (1951) or Valbuena Prat (1960), and qualified by Begoña López Bueno (2000). As for the specific case of Versos, traditionally considered closer to the Baroque, its poems are located in the network within this intermediate community, and close to Herrera. Alongside them, we find some of the poets which –according to López Bueno (2000)– make up the Sevillian transitional poetic group –e.g. Arguijo, Pacheco or Jáuregui– and the authors who write during the transitional period (López Bueno, 2006), as Cervantes, Luis Carrillo y Sotomayor or Pedro Espinosa.

**Evaluation of Literary Style Evolution. Is it Linguistic Change?**

However, there is a question that emerges: are we seeing literary style or linguistic change? To answer this question, it is necessary to verify that the Modularity communities reflect an evolution of poetic style, and not the linguistic change throughout the sixteenth and seventeenth centuries. With this aim, another analysis was performed using a non-fictional, non-literary corpus of the same time. For this purpose, the CHARTA corpus was selected (CHARTA (Corpus Hispánico y Americano En La Red: Textos Antiguos), n.d.), as it was the best proxy that could be found. Admittedly, this is not the best corpus for our research question, as it contains administrative texts. In spite of this, a better one could not be found, it is easily accessible and offers an enough quantity of texts for the period of interest in this study. Hence, those texts belonging to the dates studied and with a minimum length of 1800 words (as with the poets) have been collected from the CHARTA website. The consensus tree of this corpus was generated using the same parameters as for the consensus tree produced for the Early Modern poetry. The consensus tree CSV was imported into Gephi, resulting in a network.
to which the “Force Atlas 2” algorithm was applied. As noted, the process of creation of the consensus tree and the network was exactly the same and using the same parameters as with Early Modern poetry. As a result, the only difference between the two networks is the corpus used. CHARTA network is shown in Figs. 6 and 7.

In Fig. 6, CHARTA network displays the concrete dates of the texts and has been coloured according to that information. Lighter colours represent earlier texts, while darker colours represent later ones. Even at a first glance, it is easy to notice big differences between the CHARTA network for 16-seventeenth centuries’ texts and the Early Modern Spanish poetry network. On top of this, compared to this last one, in CHARTA network the chronological signal is not as clear in the clustering and relations between texts. This suggests that chronology has a less important role in the formation of the CHARTA network.
Figure 7 shows the communities obtained after applying the Modularity function. The result suggests that the texts were placed on the network depending not exactly on the repository where they are housed, but on the CHARTA
subcorpus of provenance (e.g. the purple community mostly includes texts from the COREECOM subcorpus, while the light blue community only includes texts from the CODEA subcorpus).

A possible explanation may be that thematic differences between CHARTA subcorpus had a strongest influence in this corpus than linguistic variation over time. For instance, the purple community includes texts from the Archivo de Indias, which are all related with America, and two particular texts from the Mexican Archivo General de la Nación dealing with similar American and colonial issues.

In any case, it seems clear of the comparison between the poetry and the non-fictional network that linguistic change and literary style are not following the same path in the Early Modern period. This reinforces that Modularity communities in Early Modern Spanish poetry are related to the stylistic change from the Renaissance to the Baroque.

**Conclusions and Future Work**

In this paper, an exploration of literary style in Early Modern Spanish poetry through Stylometry and Network Analysis has been conducted. The aim was to assess if these methodologies could shed new light on previous research, as well as on the role played by Fernando de Herrera and the poems of his posthumous edition in the change from Renaissance to Baroque.
The obtained network proves that there is a computationally measurable chronological evolution in the texts of this period, as well as a change of style from Renaissance to Baroque, with an intermediate group of transitional poets, supporting López Bueno’s theory, also measured through Modularity and Network Analysis. Both findings have been evaluated through different statistical tests. Moreover, the generation of another network with non-fiction Early Modern Spanish texts and using the same parameters suggests that the modularity communities detected in the poetic network do not correlate with language change, but with poetic style evolution.

Regarding Herrera’s role, both his undoubted poems and the new ones published in 1619 fall within the transition group between the Renaissance and Baroque, along with other authors and texts. This agrees to the vision of Herrera as a transitional poet between the two stylistic movements by literary scholars. Versos, which has been considered more Baroque, is placed on this transition community, and very close to the rest of Herrera’s poems.

This paper shows that the combination of Literature Studies with Digital Humanities and quantitative techniques—namely, Stylometry and Network Analysis—is a promising one. This is just a small sample of what may be achieved through the collaboration between these disciplines, and the new paths they can open for Spanish literature research. As future work, it would be interesting to apply other quantitative and stylometric techniques to the research question addressed here and especially to Herrera’s poems in order to explore further if Versos is actually more Baroque if compared with the rest of Herrera’s poetry, as scholars suggest. Finally, in a broader sense, other valuable future work may be obtained through the application of these methodologies to other periods and genres of Spanish literature.

Appendix

See complementary materials at this GitHub repository:

“https://github.com/lamusadecima/Network_for_Golden_Age_Spanish_Poetry.”

Acknowledgements I thank my PhD supervisor, Juan Montero (University of Seville), and the supervisors of my research stays, Maciej Eder (Computational Stylistics Group, Institute of Polish Language) and Fotis Jannidis (University of Würzburg), for their guidance, inspiring comments and suggestions on this project. I also acknowledge my colleagues José Calvo (State and University Library of Göttingen), Jaime Galbarro (University of Seville) and Joanna Byszuk (Computational Stylistics Group, Institute of Polish Language), for their valuable help with statistical techniques, historical insights and the functioning of modularity. In addition, I am grateful to the anonymous reviewers for many valuable remarks. Finally, I would like to thank Enrique and Anjuli for proofreading the final text.

Funding This article is part of Laura Hernández-Lorenzo’s PhD dissertation, “Los textos poéticos de Fernando de Herrera: aproximaciones desde la Estilística de corpus y la Estilometría”, supervised by Juan Montero at the University of Seville, and funded by a FPU (‘Formación de Profesorado Universitario’) Spanish Research and Teaching Fellowship from the Spanish Ministry of Education (FPU14/05934). The research presented in this paper also benefited from a research stay with the Computational Stylistics Group, at the Institute of Polish Language, supervised by Maciej Eder and funded by the University of Seville (VIPPT-2018-L.3), and from a research stay at the University of Würzburg, supervised by Fotis Jannidis and supported by an Erasmus Traineeship Fellowship. Finally, last stages of this study
were carried out within the projects *Hacia la Institucionalización Literaria: Polémicas y Debates Historiográficos* (1500-1844) (RTI2018-095664-B-C22), funded by the Spanish Ministry of Science (*Ministerio de Ciencia, Innovación y Universidades*), and Large-Scale Text Analysis and Methodological Foundations of Computational Stylistics (SONATA-BIS 2017/26/E/HIS2/01019) funded by the Polish National Science Center (NCN).

**Availability of Data and Material** Complementary materials for this paper, such as the corpus used or the developed code are available at the GitHub repository included in the “Appendix”.

**Code Availability** The R code used in this paper is available, along with other complementary materials at the paper’s GitHub repository, included in the “Appendix”.

**Declarations**

**Conflicts of interest** The author has no conflicts of interest or competing interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit [http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

**References**

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Third International AAAI Conference on Weblogs and Social Media*, 361–362.

Blanco, M. (2016). *Góngora o la invención de una lengua*. Universidad de León, Área de Publicaciones.

Blasco, J. (2016). Avellaneda desde la estilometría. En P. Ruiz Pérez (Ed.), *Cervantes. Los viajes y los días* (pp. 97–116). Prosa Barroca y SIAL Ediciones.

Battaglia, S. (1954). Per il testo di Fernando de Herrera. *Filologia Romana*, I, 51–88.

Blecua, J. M. (1958). De nuevo sobre los textos poéticos de Herrera. *Boletín de La Real Academia Española*, XXXVIII, 377–408.

Burrows, J. (2002). ‘Delta’: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3), 267–287.

Burrows, J. (2004). Textual analysis. In R. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A companion to digital humanities* (pp. 323–347). Blackwell.

Calvo Tello, J. (2019). Delta Inside Valle-Inclán: stylometric classification of periods and groups of his novels. *Romanistische Studien. Theorien Von Autorschaft Und Stil in Bewegung: Stilistik Und Stilometrie in Der Romania*, 6, 151–164.

Calvo Tello, J., & Cerezo Soler, J. (2018). La conquista de Jerusalén ¿de Cervantes? Análisis estilométrico sobre autoría en el teatro del Siglo de Oro español. *Digital Humanities Quarterly*, 12(1), 1–10. *CHARTA (Corpus Hispánico y Americano en la Red: Textos Antiguos)*. (n.d.). Retrieved November 2, 2021, from [http://www.corpuscharta.es](http://www.corpuscharta.es).

Cuevas, C. (1985). La cuestión textual. En F. de Herrera (Ed.), *Poesía castellana original completa* (pp. 87–99). Cátedra.

de la Rosa, J., & Súarez, J. L. (2016). The life of Lazarillo de Tormes and of his machine learning adversaries. Non-traditional authorship attribution techniques in the context of the Lazarillo. *Lemir*, 20, 373–438.
2019. Conference abstracts. https://dev.clariah.nl/files/dh2019/boa/0981.html. Accessed 30 Dec 2020.
Orozco Díaz, E. (1971). Estructura manierista y estructura barroca en la poesía. Introducción y comentar-
ios a unos sonetos de Góngora. In Historia y estructura de la obra literaria: Coloquios celebrados
del 28 al 31 de marzo de 1967 (pp. 97–116). Madrid: CSIC.
Orozco Díaz, E. (1981). Manierismo y Barroco. Cátedra.
Pepe Sarno, I. (1982). Se non Herrera, chi? Varianti e metamorfosi nei sonetti di Fernando de Herrera. I
Parte. Studi Ispanici, 33–69.
Rißler-Pipka, N. (2016a). Digital Humanities und die Romanische Literaturwissenschaft. Der
Autorschaftsstreit um den Lazarillo de Tormes. Romanische Forschungen, 128(3), 316–342.
Rißler-Pipka, N. (2016b). Avellaneda y los problemas de la identificación del autor. Propuestas para una
investigación con nuevas herramientas digitales. En H. Ehrlicher (Ed.), El otro Don Quijote. La con-
tinuación de Fernández de Avellaneda y sus efectos (pp. 27–51), Augsburg: Universität Augsburg.
Rißler-Pipka, N. (2019). In search of a new language. Measuring style of Góngora and Picasso. Roma-
nische Studien, Theorien Von Autorschaft Und Stil in Bewegung: Stilistik Und Stilometrie in Der
Romania, 6, 117–150.
Rochat, Y. (2009). Closeness Centrality Extended To Unconnected Graphs: The Harmonic Centrality
Index. Conference paper at ASNA 2009. Published by EPFL. https://infoscience.epfl.ch/record/
200525/files/%5BEN/%5DASN A09.pdf. Accessed 30 December 2020.
Rojas Castro, A. (2017). Luis de Góngora y la fábula mitológica del Siglo de Oro: Clasificación de textos
y análisis léxico con métodos informáticos. Studia Aurea, 11, 111–142.
Rotari, G., Jander, M., & Rybicki, J. (2020). The Grimm brothers: A stylometric network analysis. Digital
Scholarship in the Humanities. https://doi.org/10.1093/llc/fqz088
Ruestes, M. T. (1986). Introducción. In F. de Herrera, Poesía (p. IX–LXIV). Planeta.
Ruiz Pérez, P. (2000). Renovación del orden genérico: las Flores de poetas ilustres (1605). Calíope. Jour-
nal of the Society for Renaissance and Baroque Hispanic Potery, 9(1), 5–33.
Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2004). A companion to digital humanities.
Blackwell.
Smith, P. W. H., & Aldridge, W. (2011). Improving authorship attribution: Optimizing Burrows’ Delta
method. Journal of Quantitative Linguistics, 18(1), 63–88.
Ulla Lorenzo, A., Martínez Carro, E., & Calvo Tello, J. (2020). Las comedias de dudosa atribución
de Agustín Moreto: Nuevas perspectivas estilométricas. Neophilologus. https://doi.org/10.1007/
s11061-020-09649-3
Valbuena Prat, Á. (1960). Historia de la literatura española: (Vol. I). Gustavo Gili.
Vilanova, A. (1951). Fernando de Herrera. In Historia General de las Literaturas Hispánicas (II, pp.
689–751). Barna.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published
maps and institutional affiliations.