

Los sonetos de autoría femenina en el Siglo de Oro: aproximación estilométrica*

Sonnets written by women in the Spanish Golden Age: a stylometric approach

LAURA HERNÁNDEZ-LORENZO

Universidad de Sevilla. Facultad de Filología. C. Palos de la Frontera, s/n, 41004 Sevilla (España).

Dirección de correo electrónico: lhernandez1@us.es.

ORCID: <https://orcid.org/0000-0003-3489-2193>.

Recibido/Received: 16-1-2025. Aceptado/Accepted: 22-3-2025.

Cómo citar/How to cite: Hernández-Lorenzo, Laura (2025). “Los sonetos de autoría femenina en el Siglo de Oro: aproximación estilométrica”. *Castilla. Estudios de Literatura*, 16, pp. 175-198. DOI: <https://doi.org/10.24179/cel.16.2025.175-198>.

Artículo de acceso abierto distribuido bajo una [Licencia Creative Commons Atribución 4.0 Internacional \(CC-BY 4.0\)](#). / Open access article under a [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](#).

Resumen: En este estudio se realiza una primera aproximación cuantitativa a la escritura poética de autoría femenina en el Siglo de Oro mediante la aplicación de metodologías computacionales y estilométricas, siguiendo la estela de trabajos internacionales de gran relevancia en el campo de la Estilística computacional. Los resultados obtenidos apuntan a que es posible diferenciar la escritura femenina de la masculina en este género y periodo, y se ha podido detectar un motivo metapoético en la poesía de estas autoras relacionado con la escritura femenina.

Palabras clave: estilometría; escritoras; poesía; Siglo de Oro; topic modelling.

Abstract: This study presents a first quantitative approach to the poetic writing of female authorship in the Golden Age through the application of computational and stylometric methodologies, following the trail of international relevant works in the field of Computational Stylistics. The results obtained suggest that it is possible to differentiate between female and male writing in this genre and period. Additionally, a metapoetic motif in women's poetry related to female writing was detected.

Keywords: stylometry; women writers; poetry; spanish Golden Age; topic modelling.

* Este trabajo se ha realizado en el marco de una ayuda Juan de la Cierva (FJC2021-047998-I), financiada por MCIN/AEI/10.13039/501100011033 y por la Unión Europea «NextGenerationEU»/PRTR.

INTRODUCCIÓN

La emergencia de las Humanidades Digitales ha supuesto la aparición de nuevas metodologías y perspectivas en el estudio de la Literatura. Dentro de estas, destacan los enfoques cuantitativos (Hoover, 2013; Jannidis et al., 2017), que se han visto reforzados por las teorías de análisis distante (Moretti, 2013), macroanálisis (Jockers, 2013) y analítica cultural (Piper, 2016), y se han recogido dentro de los enfoques de minería de textos (Jockers y Underwood, 2016). A pesar de que en los últimos años han aumentado los trabajos que analizan cuantitativamente los textos de escritores en el ámbito hispánico, con nuevas aportaciones en cuestiones de autoría (Blasco y Ruiz Urbón, 2023; Bazzaco, 2021; Cuéllar y Vega García-Luengos, 2023; García-Reidy, 2019; Fradejas Rueda, 2016; Kroll y Sanz-Lázaro, 2022), movimientos literarios (Hernández-Lorenzo, 2022) o géneros textuales (Calvo Tello, 2021), se ha dedicado una menor atención a las autoras.

Con el fin de arrojar nueva luz en la aplicación de estos métodos a la escritura femenina, en este trabajo se presentan resultados de una investigación en marcha, centrada en el análisis cuantitativo de los textos poéticos de las escritoras del Siglo de Oro. El principal objetivo es explorar si las técnicas estilométricas pueden diferenciar la escritura femenina de la masculina y realizar una primera aproximación a las características de la poesía escrita por mujeres en el periodo áureo.

La estructura de este trabajo es la siguiente: después de esta introducción, se presenta un breve estado de la cuestión de los estudios cuantitativos aplicados a la literatura escrita por mujeres. A continuación, se describe la preparación del corpus utilizado en el estudio y se exponen la metodología y herramientas empleadas. Seguidamente, se presentan los resultados de los análisis llevados a cabo. Por último, se exponen las conclusiones alcanzadas.

1. ESTUDIOS CUANTITATIVOS APLICADOS A LA LITERATURA ESCRITA POR MUJERES

Una de las disciplinas que suele incluirse dentro de los enfoques cuantitativos es la estilometría. Aunque desde sus orígenes ha predominado su aplicación a problemas de atribución de autoría, en los últimos años han aumentado significativamente las aplicaciones

estilométricas para analizar otras cuestiones estilísticas, en lo que se ha llamado la estilometría más allá de la autoría.

Dentro de esta vertiente, contamos con estudios previos que han empleado técnicas computacionales y estilométricas para aproximarse a las diferencias entre el discurso masculino y el femenino, aunque fundamentalmente aplicados a textos en inglés. Entre los primeros precedentes, cabe destacar “Automatically Categorizing Written Texts by Author Gender” (Koppel et al., 2002), en el que se consigue detectar de forma automática el sexo del autor de un texto con un índice de acierto del 80 %; y “Gender, genre, and writing style in formal written texts” (Argamon et al., 2003), en el que ya se señalan algunos rasgos distintivos, como el mayor uso de pronombres por parte de las mujeres. Se alcanzan conclusiones similares en *The Secret Life of Pronouns* (Pennebaker, 2011), que cuenta con un capítulo dedicado a esta cuestión. En *Macroanalysis* (Jockers, 2013) encontramos también un capítulo dedicado a analizar las diferencias entre la escritura masculina y la femenina, esta vez aplicado a novelas inglesas del XIX, con interesantes resultados, y en la enciclopedia *Literary Studies in the Digital Age*, editada por la MLA, figura un capítulo titulado “Textual Analysis” (Hoover, 2013), en el que se presenta una lista de las cien palabras más características de poetas masculinos y femeninos en lengua inglesa. En 2016, Rybicki dedica uno de sus trabajos, realizado en el marco de la COST-Action “Women Writers in History: Toward a New Understanding of European Literary Culture”, a analizar un corpus de novela inglesa del XIX de autoría femenina para detectar si alguna de las novelas incluidas ha sido escrita en realidad por un hombre (Rybicki, 2016), detectando con estos métodos dos casos altamente probables.

En 2018, se publican tres estudios relacionados con esta cuestión: el primero (Savoy, 2018) es un análisis de autoría de las novelas de Elena Ferrante para corroborar si detrás de esta escritora se encuentra un autor masculino, Starnone. En el segundo (Weidman y O’Sullivan, 2018), se exploran los marcadores estilísticos masculinos y femeninos en corpus narrativos de tres periodos distintos (época victoriana, modernismo y época contemporánea) a través de métodos estilométricos de distinción léxica. El tercero es la monografía *Reading Beyond the Female* (Koolen, 2018). En esta última obra, para el caso de la novela holandesa actual, se

realiza un análisis de la relación entre el sexo del autor y la percepción de la calidad literaria utilizando métodos computacionales.¹

Los resultados de interés obtenidos por los trabajos anteriores suponen un aliciente para aplicar estos métodos a la literatura en español. Sin embargo, en el ámbito de la literatura hispánica, únicamente contamos con un estudio reciente sobre las diferencias sociolingüísticas de género en la novela del siglo XIX (Bermúdez-Sabel, 2023), y son escasos los trabajos que analizan cuantitativamente el texto de las obras literarias de las escritoras. Para el caso de las escritoras españolas de la primera Edad Moderna, destacan los esfuerzos llevados a cabo en el proyecto BIESES (Baranda et al., 2019), y algún estudio estilométrico de obras atribuidas a escritoras (Hernández-Lorenzo y Byszuk, 2023), pero lo cierto es que se ha prestado una menor atención a las autoras, que además se encuentran infrarrepresentadas en los corpus utilizados debido a varias causas: fundamentalmente, una producción femenina significativamente menor que la masculina, su limitada inclusión en el canon, las dificultades para recopilar sus textos y la escasa digitalización de los mismos.

Por estas razones, el presente estudio aspira a contribuir al estado de la cuestión de los estudios cuantitativos de la escritura femenina hispánica en el ámbito de la primera Edad Moderna mediante un análisis computacional de textos poéticos de escritoras del Siglo de Oro.

2. PREPARACIÓN DEL CORPUS

Para poder llevar a cabo este estudio, ha sido necesaria la preparación de un corpus de sonetos de escritoras del Siglo de Oro, cuyos textos han sido extraídos de diversas fuentes, que van desde manuscritos e impresos hasta ediciones y antologías modernas (Baranda, 2019; Carvajal y Mendoza, 1990; Fernández López, 2015; Gorría, 2018; Marcos, 2022; Olivares y Boyle, 2011; Ramírez de Guzmán, 2004; Zayas, 2014, 2022).²

¹ La metodología incluye el análisis computacional de novelas escritas por hombres y mujeres, así como encuestas sobre la calidad de las obras, y el objetivo es investigar si existen rasgos en la escritura femenina que justifiquen que sus obras sean generalmente percibidas como de menor calidad. Es decir, se trata de ver hasta qué punto el sexo del autor se corresponde con los textos que escribe o hay otras variables que influyen, y en este sentido, se alerta sobre los sesgos producidos por los estereotipos.

² En cuanto a los manuscritos e impresos utilizados, en el caso de María de Zayas, se han recogido sus sonetos insertos como poemas preliminares en las obras de otros autores, consultadas a través de la Biblioteca Digital Hispánica de la BNE; las obras poéticas de

Para asegurar que no hubiera diferencias gráficas debidas a los criterios ortográficos de las ulteriores fuentes y ediciones (las cuales podrían interferir en los resultados), la ortografía de todos los poemas se ha revisado, modernizado y normalizado a las reglas del español actual.

En este momento, el corpus comprende 74 autoras y 359 sonetos de los siglos XVI y XVII. Sin embargo, entraña una serie de limitaciones: en primer lugar, al tratarse de textos poéticos, las muestras textuales son breves; en segundo lugar, por razones históricas, de la mayoría de las autoras se nos ha transmitido un único poema, y solo se han conseguido recopilar ocho poemas o más de diez autoras, y quince poemas o más de seis escritoras (véase la Tabla 1).³ Estas limitaciones se tomarán en consideración a la hora de realizar los diferentes análisis y de interpretar los resultados obtenidos.

Autora	Fecha	Sonetos
Isabel de Vega	h. 1558-¿?	8
Luisa de Carvajal y Mendoza	1566-1614	8
Sor Cecilia del Nacimiento	1570 - 1646	9
Ana de Trinidad	1577-1613	19
María de Zayas	1590-1661	26
Violante do Ceu	1601-1693	86
Catalina Ramírez de Guzmán	1611-1684 o 1685	8
Leonor de la Cueva	1611-1705	21
María de Santa Isabel	1613-1665	15
Sor Juana Inés de la Cruz	1648 o 1651-1695	73

Tabla 1. Autoras del Siglo de Oro de las que se nos han transmitido al menos 8 poemas, ordenadas por fecha de nacimiento.

Violante do Ceu, las *Rimas varias* y el *Parnaso lusitano*, fueron también consultados a través de las digitalizaciones de la Biblioteca Digital Hispánica de la BNE: <http://bdh-rd.bne.es/viewer.vm?id=0000262952&page=1> y <http://bdh-rd.bne.es/viewer.vm?id=0000252982> [consulta: 16/01/2025]). Los sonetos completos de Leonor de la Cueva han sido extraídos de la edición preparada por Bárbara Vela para su Trabajo de Fin de Máster, tutorizado por María D. Martos, a quien agradezco que me haya dado acceso a estos textos. Para el resto de las autoras, los sonetos han sido localizados fundamentalmente a través de la base de datos de Bieses: <http://62.204.211.38:8080/bieses/> [consulta: 16/01/2025].

³ A pesar de que son muy pocas las autoras de las que se ha podido recoger una cantidad suficiente de poemas, se trata de escritoras representativas de la escritura poética en este periodo, muestra de la cual es que se incluye a todas las autoras estudiadas en la monografía sobre la escritura femenina en los sonetos áureos de Fox (2008), nómina a la que el presente estudio agrega a Isabel de Vega, Cecilia del Nacimiento, Ana de Trinidad y María de Zayas.

Con el objetivo de comparar la escritura femenina con la masculina del mismo periodo y subgénero en algunos análisis, se utilizará un corpus de sonetos de escritores (Hernández-Lorenzo, 2019), que se basa en el desarrollado por el proyecto ADSO (Navarro-Colorado et al., 2018), y del que se retiraron los sonetos de Sor Juana Inés de la Cruz (la única autora incluida) al corpus de escritoras. Este corpus es mucho mayor en tamaño y se encuentra más balanceado en tanto cantidad de textos por autor que el que se ha podido recoger de las autoras, lo cual también se toma en consideración a la hora de diseñar los análisis e interpretar los resultados.

3. METODOLOGÍA

Este estudio utiliza la metodología de los enfoques cuantitativos dentro de las humanidades digitales y de la estilometría más allá de la autoría, siguiendo los trabajos citados en el estado de la cuestión (sección 1). Se combinan métodos clásicos de esta disciplina, como los análisis de grupos, con la aplicación de medidas de distinción o *keyness*, como la aplicación de Zeta (Burrows, 2007; Craig y Kinney, 2009) para contrastar corpus. Estos análisis han sido realizados mediante el paquete *stylo* (Eder et al., 2016), que además de ser ampliamente utilizado, está bien documentado y garantiza la replicabilidad de los resultados.

El estudio se completa con la aplicación de *topic modelling* con el objetivo de realizar una aproximación a los diferentes temas que subyacen en la poesía femenina. Siguiendo los parámetros con los que se obtienen mejores resultados en un estudio previo que aplica *topic modelling* a un corpus de sonetos del Siglo de Oro (Navarro-Colorado, 2018), se ha decidido aplicar el algoritmo LDA (*Latent Dirichlet Allocation*) sobre los textos sin lematizar.⁴ Esta metodología se ha aplicado a través de la implementación del clásico algoritmo LDA en el paquete *topicmodels* (Grün y Hornik, 2011), que está bien documentado, permite trabajar con los paquetes *tidyverse*, desarrollados para realizar análisis de minería de textos de forma rápida y eficiente, y cuenta con funciones para crear gráficos y evaluar automáticamente la coherencia de los temas obtenidos.

Por último, debido a su especial interés, uno de los tópicos o temas obtenidos es analizado con mayor detenimiento a través del análisis de las palabras que lo conforman. Para ello, se ha utilizado *Voyant tools* (Sinclair

⁴ Se trata de un método de minería de textos por el que se obtienen los principales temas de un texto a partir de las palabras que tienden a aparecer en los mismos contextos.

y Rockwell, 2016), que ha permitido obtener las frecuencias de uso de estos términos en las diferentes autoras, así como los contextos concretos (concordancias) en las que los utilizan.

4. RESULTADOS⁵

4. 1. Métodos estilométricos no supervisados: análisis de grupos

En el primer análisis, se ha empleado un método estilométrico no supervisado, el análisis de grupos, para realizar una primera aproximación a la diferenciación de la poesía de autoría femenina de la masculina. Se trata de un método estadístico cuya finalidad es “agrupar una serie de elementos en grupos de manera que se dé la máxima homogeneidad posible dentro de cada grupo y, a la vez, la mayor diferencia entre los diversos grupos” (Fradejas, 2016, p. 206).

Para que pueda llevarse a cabo el análisis, se han introducido los poemas de las autoras con quince poemas o más, cercanas al límite de extensión de 2000 palabras señalado en la bibliografía (Eder, 2017; Hernández-Lorenzo, 2019).⁶ Junto a estas, se han añadido los textos de seis autores representativos del mismo periodo (en este caso, del siglo XVII): Góngora (1561-1627), Lope de Vega (1562-1635), Juan de Arguijo (1567-1623), Quevedo (1580-1645), Juan de Jáuregui (1583-1641) y Pedro Soto de Rojas (1584-1658). Se ha aplicado la medida Cosine Delta (Evert et al., 2017), por ser la medida de distancia que generalmente obtiene los mejores resultados, sobre las mil palabras más frecuentes. Como puede comprobarse en la Figura 1, se aprecia una distinción clara entre las obras de autoría femenina y las de autoría masculina, que aparecen en grupos diferentes.

⁵ Debido a las restricciones de extensión de este estudio, se incluyen los resultados obtenidos de mayor interés entre todos los análisis realizados. Los resultados completos pueden consultarse en el repositorio del trabajo: <https://github.com/lamusadecima/poetisasSO/tree/main>.

⁶ A la hora de realizar análisis cuantitativos como los estilométricos, resulta fundamental contar con una extensión significativa de texto, pues de otra forma los resultados no serán fiables. Esto se ha estudiado de forma sistemática en los estudios citados de evaluación de atribución de autoría, en los que se ha comprobado de forma experimental que es necesario contar con aproximadamente 2000 palabras para garantizar la fiabilidad de los resultados.

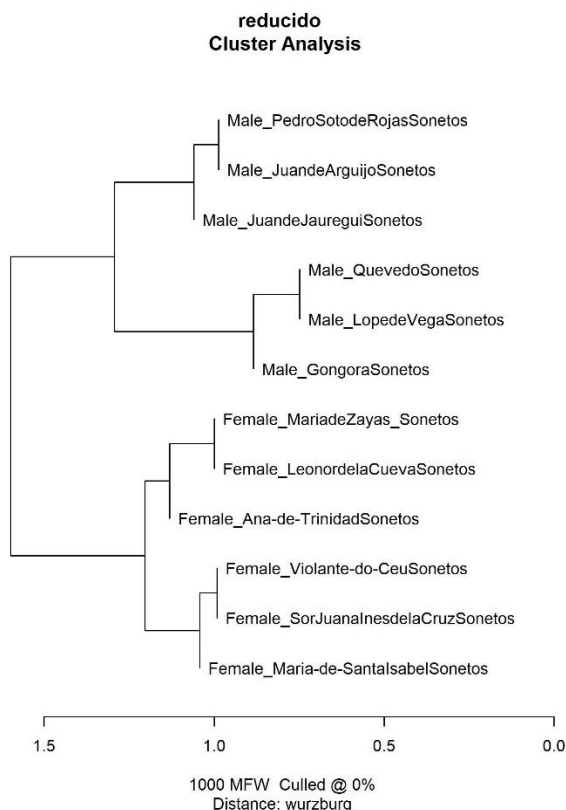


Figura 1. Análisis de grupos aplicando Cosine Delta a las 1000 palabras más frecuentes. Fuente: elaboración propia.

4. 2. Medidas estilométricas de distinción o *keyness*: Zeta

A continuación, se ha utilizado un método estilométrico de contraste, *oppose*, que permite comparar corpus a través de la medida de distinción Zeta (Eder et al., 2016). Mediante la aplicación de esta técnica, se pretende localizar qué palabras o formas son inusualmente más frecuentes en el corpus de las poetisas frente al de los poetas varones (Craig y Kinney, 2009), siguiendo la citada aplicación de este método a corpus de textos en inglés con obras de autoría masculina y femenina (Weidman y O'Sullivan, 2018). Además, para comprobar si la diferencia de tamaño entre los dos corpus comparados influye en los resultados, se ha realizado un primer análisis con los corpus completos y tres más con muestras aleatorias del

corpus de escritores con un tamaño similar al de las escritoras, obteniéndose resultados similares en todos los análisis. En la Figura 2, se representan los marcadores estilísticos de los autores varones (triángulos) frente a los de las autoras (círculos) en un gráfico bidimensional, que aparecen bien definidos, pues no se produce solapamiento.

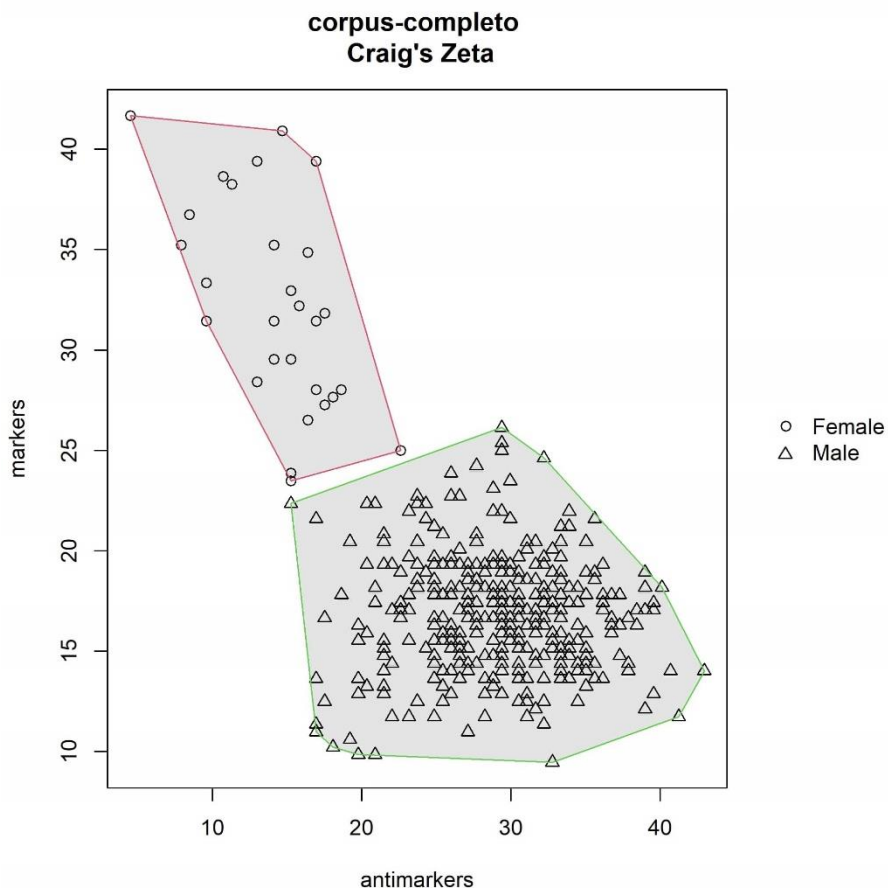


Figura 2. Gráfico bidimensional que muestra los resultados de la aplicación de *oppose* (Zeta de Craig) sobre los corpus de sonetos de escritoras (*primary set*) y escritores (*secondary set*) del Siglo de Oro. Fuente: elaboración propia.

En cuanto a los resultados léxicos, estos muestran que la poesía de las escritoras se caracteriza por el léxico religioso⁷ (por ejemplo: *soberano, deidad, divino, madre, Dios, sacro, grandeza, peregrina, querido, quererte*,

⁷ Pueden consultarse las listas completas en el repositorio del estudio: <https://github.com/lamusadecima/poetisasSO/tree/main/zeta>.

amado), mientras que en el corpus de autoría masculina emergen palabras relacionadas con el amor pasional, posesivos de primera persona del singular, palabras relacionadas con la naturaleza, el tiempo, la violencia y el materialismo (por ejemplo: *ciego, loco, lágrimas, mío, mis, mí, mía, monte, mar, viento, armas, guerra*). Ahora bien, debe tenerse en cuenta la posible influencia de algunos factores como la condición social de los escritores (con un predominio de religiosas en el corpus femenino), y especialmente la temática y el género textual, en términos de sonetos religiosos frente a otro tipo de sonetos (amorosos, satíricos...). En este sentido, Koolen (2018) advierte que el género textual interfiere a menudo en los estudios comparativos entre obras escritas por hombres y por mujeres.

4. 3. *Topic modelling*

Con el objetivo de realizar una aproximación a los principales temas que subyacen en el corpus de autoría femenina, el siguiente paso fue realizar un análisis de *topic modelling*. Se trata de un método de minería de textos por el que se obtienen los principales temas de un texto a partir de las palabras que tienden a aparecer en los mismos contextos. El algoritmo más utilizado es *Latent Dirichlet Allocation*, comúnmente conocido como LDA (Blei, 2012). Este método parte de la idea que a la hora de escribir un texto existen una serie de temas predefinidos a los que pertenecen determinadas palabras, de forma que el autor va eligiendo palabras de diferentes temas para construir su texto. Así, el algoritmo pretende reconstruir estos temas haciendo el proceso inverso.

Esta metodología se ha aplicado anteriormente con éxito a textos narrativos (Jockers, 2013; Jockers y Mimno, 2013), teatrales (Schöch, 2017) y poéticos (Rhody, 2012; Lou et al., 2015; Navarro-Colorado, 2018). Estos últimos estudios señalan cómo las características del lenguaje poético provocan que los resultados obtenidos con *topic modelling* tengan ciertas particularidades en comparación a los obtenidos con otros tipos de textos: “topic keyword distributions in poetry are not ‘thematic’ in the way that topic models of non-fiction documents are” (Rhody, 2012, p. 23). Esta investigadora señala cómo los grupos resultantes, más que temas, representan formas del discurso, y, por tanto, pueden ayudar a descubrir patrones latentes en los textos. En la misma línea, Borja Navarro concluye que *topic modelling* aplicado a poesía extrae fundamentalmente motivos

poéticos, esto es, palabras recurrentes o imágenes poéticas que aparecen juntas y que se refieren a uno o más temas (Navarro-Colorado, 2018).⁸

Esta metodología se ha aplicado al corpus de poesía áurea de autoría femenina siguiendo los parámetros más exitosos en el estudio previo citado de sonetos del Siglo de Oro (Navarro-Colorado, 2018). Puesto que no existe un número predefinido de temas para cada texto, se han entrenado modelos de 10, 15, 20, 25 y 50 temas. Tras realizar una evaluación de estos, tanto manual como automática a través de medidas de coherencia de temas (Mimno et al., 2011), los mejores resultados fueron los obtenidos con los modelos de 15 temas (véase la Tabla 2 con los valores de coherencia de temas obtenidos).

Tema	Valor con corpus completo	Valor con corpus reducido
1	-135.8490	-87.90741
2	-126.9437	-106.57250
3	-160.8560	-85.93389
4	-144.4752	-63.04488
5	-153.6702	-79.28076
6	-146.9720	-86.47301
7	-148.7625	-107.68658
8	-156.2738	-101.55814
9	-142.4196	-100.37092
10	-137.5409	-94.26401
11	-146.0330	-105.53712
12	-117.3186	-78.47551
13	-144.7284	-103.71707
14	-140.2920	-91.07809
15	-143.8448	-89.14877

Tabla 2. Valores de coherencia de temas obtenidos para 15 temas. Los números más cercanos a 0 indican mayor coherencia.

Aunque esta técnica se ha aplicado tanto sobre el corpus completo como sobre un corpus reducido con muestras aleatorias, nos centraremos

⁸ Frente a metodologías cualitativas de análisis semántico como la isotópica, *topic modelling* permite realizar un análisis estadístico de las principales estructuras semánticas de grandes cantidades de texto, por lo que resulta idóneo para el estudio cuantitativo y estilométrico que se presenta en este trabajo.

a continuación en el segundo. Esta decisión se debe a que en el corpus completo se dan grandes diferencias en el material textual disponible, ya que algunas autoras están sobrerrepresentadas, como Violante do Ceu con 86 poemas, mientras que de otras solo se ha podido recoger un poema, lo cual podría estar produciendo sesgos en los resultados, fundamentalmente relacionados con la poesía religiosa, frecuente en esta escritora.⁹ En cambio, el corpus reducido comprende únicamente las autoras con, al menos, ocho poemas (véase más arriba la Tabla 1). Además, para que ninguna autora se encuentre sobrerrepresentada, de las autoras que han escrito más de diez poemas, se han seleccionado diez de forma aleatoria. De esta forma se asegura que los resultados obtenidos son representativos de la poesía de todas las autoras incluidas y no de las que escribieron más textos. Los resultados pueden consultarse en la Figura 3 y la Tabla 3.¹⁰

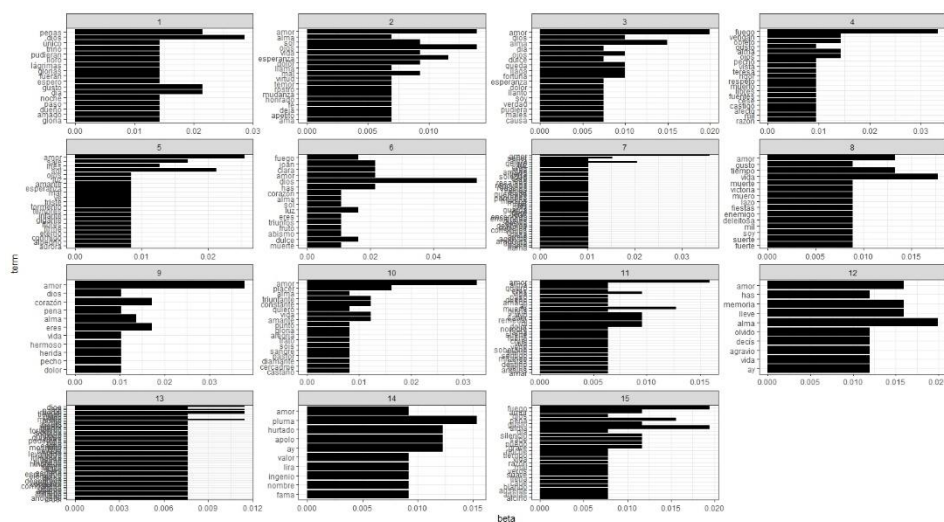


Figura 3. Resultados del análisis de *topic modelling* sobre el corpus reducido aleatorio, entrenado sobre 15 temas. Fuente: elaboración propia.

⁹ Los resultados del corpus completo pueden consultarse en el repositorio del estudio: <https://github.com/lamusadecima/poetisasSO/tree/main/TM/corpus-completo>.

¹⁰ La Figura 3 puede consultarse con mayor resolución en el repositorio del estudio: <https://github.com/lamusadecima/poetisasSO/tree/main/TM/corpus-reducido>. Como puede observarse, a pesar de haber afinado los parámetros previamente, no todos los temas obtenidos dentro de un mismo modelo tienen la misma consistencia e interpretabilidad, lo cual queda también reflejado en los diferentes valores de coherencia obtenidos en la Tabla 2. Esto es habitual en *topic modelling* y, por ello, nos centraremos a continuación en los tópicos más consistentes y que revistan mayor interés.

Tema 1	Tema 2	Tema 3	Tema 4	Tema 5
Dios día gusto penas dueño fueran lloro noche pudieran paso	amor ojos esperanza vida sol dolor mal deja alma temor	amor alma Dios fortuna llaga queda ojos esperanza día dolor	fuego vengan alma objeto ojos gusto Teresa razón mil afecto	amor sol sale Inés esperanza albedrío mal ojos aurora eterno
Tema 6	Tema 7	Tema 8	Tema 9	Tema 10
Dios Joan amor has Clara fuego dulce luz alma sol	amor quiero Señor vida alma eres ley aqueste luz casa	vida amor tiempo gusto muerte victoria fiestas mil deleitosa muero	amor eres corazón alma Dios vida pena dolor hermoso herida	amor placer vida amante constante triumfante alma quiero castaño sois
Tema 11	Tema 12	Tema 13	Tema 14	Tema 15
amor muerte vida vivo dolor remedio di saber alma quiero	alma amor lleve memoria vida has olvido decís ay agravio	Dios tiempo punto mortal Parca vida sol muerte dulce digna	pluma hurtado ay Apolo amor valor fama nombre ingenio lira	fuego alma pena amor gusto sabe silencio puedo grave Dios

Tabla 3. Resultados del análisis de *topic modelling* con el corpus reducido aleatorio de autoría femenina: las diez primeras palabras de cada tema.

Si se observan los resultados anteriores, resulta llamativo cómo al restringir el análisis al corpus reducido aleatorio dejan de predominar los asuntos religiosos y aparecen varios relacionados con el amor profano: los temas 2, 8 y 10 parecen estar más relacionados con la poesía amorosa, mientras que los temas 3, 7 y 9 son de contenido religioso. A pesar de la particularidad de las aplicaciones de *topic modelling* a textos poéticos, en la que, debido a la naturaleza del lenguaje poético, es más difícil establecer una correspondencia de los temas obtenidos con unidades temáticas (Rhody, 2012; Navarro-Colorado, 2018), y de la menor extensión del corpus empleado, algunos de los temas resultantes resultan de especial interés. Este es el caso del tema 14, que parece estar relacionado con la escritura, pues incluye en sus primeros puestos *pluma*, *Apolo*, *fama*, *nombre*, *ingenio*, *lira*, *valor* o *musa*.

4. 4. Examen de un tema particular: la escritura

Por su interés, a continuación, se examina con mayor detenimiento el tema 14, relacionado con la escritura femenina y que parece tener un carácter metapoético. En este sentido, especialistas como María Martos ya han señalado cómo se va imponiendo entre las poetisas una voluntad de autorrepresentación en el texto y una conciencia autorial, así como el interés de profundizar en los textos de contenido metapoético en los que las escritoras se refieren a su propia labor creadora (Martos, 2018).

Para profundizar en este tema metapoético, el primer paso ha sido explorar el uso de las palabras que lo conforman en las escritoras analizadas. Se han extraído las frecuencias absolutas y relativas a través de la herramienta *Voyant tools* (Sinclair y Rockwell, 2016).

En primer lugar, se ha examinado el lema *pluma*, que está presente en los poemas de siete de las diez autoras. Aunque en los sonetos de Violante do Ceu encontramos el mayor número de ocurrencias (4 en total), si tomamos en cuenta la extensión de los textos, la autora que más lo utiliza es María de Santa Isabel (con una frecuencia de 0,02 %). Además, si observamos los pasajes concretos en los que aparece este término, queda de manifiesto cómo en las diferentes autoras se convierte en una metáfora de la propia escritura:

1. “Si llegara mi **pluma**, oh gran Hurtado” (Isabel de Vega)
2. “Si la **pluma** de Ovidio mereciera” (María de Santa Isabel)
3. “Si mi llanto a mi **pluma** no estorbara” (María de Zayas)

4. “Mas, **pluma** mía, para” (María de Zayas)
5. “[...] y, hasta estos rasgos mal formados, sean / lágrimas negras de mi **pluma** triste” (Sor Juana)

Cabe destacar cómo este término aparece junto al tema de la *humilitas*. Además, algunos de los casos presentados pertenecen a la poesía elegíaca, como ocurre con los de María de Zayas y Sor Juana. Resulta también especialmente interesante cómo Violante do Ceu usa este término como expresión de autoconciencia y autoafirmación: “Yo tomaré la pluma y de tus glorias”. En otras ocasiones, adquiere tintes de humildad al referirse a la propia escritura, a la que califica como “pluma indigna” o “tosca pluma”.

En contraste con pluma, *lira* apenas lo usan cuatro autoras: María de Zayas, Leonor de la Cueva, María de Santa Isabel y Sor Juana Inés de la Cruz. La mayor representatividad de este lema lo encontramos en los sonetos de María de Zayas (con una frecuencia de 0,008 ‰). Como puede observarse en los siguientes ejemplos, se trata de otro término clave para referirse a la escritura, pero aporta el matiz de creación poética:

1. “y si Apolo su **lira** me prestara” (María de Santa Isabel)
2. “¡oh Fénix de la patria, oh nuevo Apolo!, / de mi **lira** te hiciera un mausoleo” (María de Zayas)
3. “mi frágil **lira** de inmortal presuma” (María de Zayas)
4. “Perdona a mi zampoña licenciosa, / si el escuchar tu **lira** delicada” (Sor Juana)

Como puede observarse, *lira* aparece acompañada de otras palabras del tema metapoético, como *Apolo*, el dios de las artes, o de expresiones que apuntan a la perdurabilidad de la palabra poética, como en el caso de los versos de María de Zayas. En el caso de Sor Juana, se compara la *lira* con la *zampoña*, y utiliza esta segunda para rebajar el valor de su propia poesía frente a la lira delicada del músico al que elogia.

Ingenio, en cambio, es utilizado por la mayoría de las autoras, siendo significativo su alto uso en los sonetos de María de Santa Isabel (con una frecuencia relativa de 0,04 ‰). En la obra de esta escritora, este término aparece relacionado con el estilo laudatorio:

1. “Alabaros, don Pedro, no es mi intento, / ni el **ingenio** a la pluma ha concedido / saber encarecer el que ha excedido / a todo estilo de encarecimiento” (María de Santa Isabel)
2. “Que sois de un **ingenio** peregrino, / de una mujer prodigio entre mortales” (María de Santa Isabel)

Nótese como en el primer fragmento, *ingenio* aparece junto a *pluma*, otra de las palabras de este tema metapoético. Cabe también señalar que, en el segundo ejemplo, no se refiere al propio ingenio, sino que se alaba el de otra mujer escritora. Destaca también el uso de este término por Violante do Ceu, cuyos versos hacen referencia de forma clara a un ingenio femenino, acompañado del tema de la *humilitas*, junto a la osadía, y en ocasiones también por otros términos del tema metapoético, como *voz*:

1. “Grande osadía es, que a lo divino / elija por asunto humano canto, / pero mayor, Señor, que aspire a tanto / la humildad de un **ingenio** femenino.” (Violante do Ceu)
2. “y expóngase al rigor de la censura / el **ingenio**, la **voz** y la osadía” (Violante do Ceu)

En el caso del lema *musa*, destaca su uso en la poesía de Isabel de Vega (con una frecuencia relativa de 0,04 ‰). A pesar de que se nos han transmitido pocos poemas de esta autora, este término se repite en su obra, por lo que resulta más representativo. Seguramente su alto uso se debe a que esta escritora utiliza frecuentemente en su poesía una voz masculina (Baranda, 2019; Martos, 2018), por lo que su uso respondería a la habitual concepción masculina de un poeta varón que crea y una musa mujer. Sin embargo, destacan algunos casos en los que Isabel de Vega expresa su identidad, identificándose como hablante mujer, como el incluido en la siguiente relación, junto a otros casos relevantes en otras autoras, donde *musa* aparece relacionado con la *humilitas* y con la autoría femenina:

1. “mi baja **musa** hace perder de vuelo” (Ana de Trinidad)
2. “Mi **musa** con estilo no hurtado / [...] mi **musa** indigna de tan gran renombre” (Isabel de Vega)¹¹

¹¹ Nieves Baranda señala cómo en estos versos muestran “una clara conciencia autorial al ofrecerse a usar su arte para consolar al poeta, en una identificación implícita con Orfeo” (Baranda, 2019, p. 108).

3. “Y pues honrar debéis lo femenino / dad a **musa** incapaz, a voz impura, / si no dulce favor, perdón benigno” (Violante do Ceu)

Nótese, además, cómo en el ejemplo de Violante do Ceu, *musa* aparece junto a *voz*, otro de los términos del tema metapoético.

Por último, dos términos de este tema que en ocasiones se refieren también al ejercicio de la escritura son *voz* y *memoria*. Para ambos, encontramos casos de especial interés en la poesía de Violante do Ceu:

1. “Mas si entre los que el mundo te dedica, / oh Padre celestial, oh sacro Atlante, / puede tener mi **voz** tan alto empleo, / yo cantaré la gloria dominica, / yo mostraré también, aunque ignorante, / que aplauso de tu luz es mi deseo”
2. “Pero tanta piedad ejecutada, / tan contino favor, tanta defensa, / tal sufrimiento de una y otra ofensa, / ¿cómo podrá cantar **voz** limitada?”
3. “[...] de Vos y de este imperio sacrosanto / canta mi ronca **voz** en verlo indigno”
4. “Por esta alcanzaré, dichosa parte, / la **memoria** inmortal que solicito / de tan divino asunto en la grandeza”

En los versos anteriores pueden apreciarse referencias a su propia voz autorial, de nuevo vinculada a la *humilitas* (“ronca voz”, “voz limitada”), junto con el deseo de que su poesía trascienda (“la memoria inmortal que solicito”).

CONCLUSIONES

En este estudio se ha realizado una primera aproximación al análisis cuantitativo de la poesía del Siglo de Oro de autoría femenina a través de los métodos de la estilometría y el *topic modelling*. A pesar de las limitaciones del corpus y de los métodos, cabe destacar cómo en un primer ensayo los resultados apuntan a que las técnicas estilométricas pueden diferenciar la escritura femenina de la masculina en la poesía del Siglo de Oro, lo que apunta a la existencia de una señal de género medible computacionalmente.

En segundo lugar, los resultados del contraste léxico apuntan a algunas diferencias de vocabulario y, especialmente, a la mayor representatividad de los términos religiosos en la poesía de autoría

femenina. Para explorar si se trata verdaderamente de un tema o un conjunto de temas especialmente prominente en este corpus, se ha realizado a continuación un análisis de *topic modelling*, cuyos resultados ponen en evidencia la importancia de la preparación del corpus y de que este sea lo más equilibrado posible, ya que el análisis con el corpus completo muestra un predominio de temas religiosos, pero, al aplicarlo sobre el corpus reducido con muestras aleatorias, los motivos religiosos se reducen de forma muy significativa y aparecen otros relacionados con el amor humano. Un tema de especial interés y que se mantiene en ambos análisis es el relacionado con la escritura, de contenido metapoético. El análisis detenido del uso de las palabras más importantes de este tema en las autoras más representativas ha permitido profundizar en cómo se refieren a la propia escritura y a la escritura femenina entre los polos de la humildad y la autoafirmación, en línea con lo que han señalado otros estudiosos mediante métodos no computacionales (Martos, 2018; Olivares y Boyce, 2021), pero añadiendo nuevos datos y matices obtenidos gracias al enfoque cuantitativo y estadístico de las herramientas digitales, que proporcionan una perspectiva complementaria.

Ahora bien, este es un trabajo en progreso y, si pudiera ampliarse el corpus de las autoras recogidas mediante la aparición de nuevos sonetos, se podrían conseguir resultados más sólidos. Especialmente como trabajo futuro, se plantea el interés de estudiar elementos más allá del léxico, como la sintaxis o la versificación, de forma que pueda obtenerse una imagen más completa de las características estilísticas que presenta la autoría femenina en los sonetos de nuestro Siglo de Oro.

BIBLIOGRAFÍA

- Argamon, Shlomo, et al. (2003). "Gender, Genre, and Writing Style in Formal Written Texts". *Text & Talk*, 23. 3, pp. 321-46.
- Baranda Leturio, Nieves (2014). "Isabel de Vega, poeta con musa (Alcalá, 1558, 1568)". *Epos*, 30, pp. 99-112.
- Baranda, Nieves, et al. (2019) "BIESES. Escritoras de la Edad Moderna, desde la bibliografía a las redes". En Leticia Sánchez Hernández (ed.), *Mujeres en la Corte de los Austrias: una red social, cultural, religiosa y política*. Madrid: Polifemo, pp. 55-82.

- Bazzaco, Stefano (2021). “Experimentos de estilometría en el ámbito de los libros de caballerías. El caso de atribución de un original italiano: Il terzo libro di Palmerino d’Inghilterra (Portonari, 1559)”. En Meritxell Simó *et al.* (eds.). “*Prenga xascú ço qui millor li és de mon dit*”. *Creació, recepció i representació de la literatura medieval*. San Millán de la Cogolla: Cilengua. Centro Internacional de Investigación de la Lengua Española, pp. 149-166.
- Bermúdez Sabel, Helena (2023). “Exploring genderlect markers in a corpus of Nineteenth century Spanish novels”. En Anne Baillot *et al.* (eds.). *Digital Humanities 2023: Book of Abstracts*. Graz: ADHO, pp. 121-123. DOI: <https://doi.org/10.5281/zenodo.7961822>.
- Blasco Pascual, Francisco Javier y Ruiz Urbón, Cristina (2023). *Análisis de textos desde la Estilometría*. Salamanca: Ediciones Universidad de Salamanca.
- Blei, David M. (2012). “Probabilistic Topic Models”. *Communications of the ACM*, 55, pp. 77-84.
- Burrows, John (2007). “All the Way Through: Testing for Authorship in Different Frequency Strata”. *Literary and Linguistic Computing*, 22. 1, pp. 27-47.
- Calvo Tello, José (2021). *The Novel in the Spanish Silver Age. A Digital Analysis of Genre using Machine Learning*. Bielefeld University Press.
- Carvajal y Mendoza, Luisa de (1990). *Poesías completas*. María Luisa García-Nieto Onrrubia (ed.). Badajoz: Diputación Provincial de Badajoz.
- Craig, Hugh y Kinney, Arthur F. (eds) (2009). *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press.
- Cuéllar, Álvaro y Vega García-Luengos, Germán (2023). “*La francesa Laura*. El hallazgo de una nueva comedia del Lope de Vega último”.

- Anuario Lope de Vega Texto literatura cultura*, 29, pp. 131-98. DOI: <https://doi.org/10.5565/rev/anuariolopedevega.492>.
- Eder, Maciej (2017). "Short samples in authorship attribution: a new approach". *Digital Humanities 2017. Conference abstracts*. McGill University & Université de Montréal. August 8-11, 2017. Montréal: ADHO, <https://dh2017.adho.org/abstracts/341/341.pdf> [16/01/2025].
- Eder, Maciej et al. (2016). "Stylometry with R: A Package for Computational Text Analysis". *The R Journal*, 8. 1, pp. 107-121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html> [16/01/2025].
- Evert, Stefan, et al. (2017). "Understanding and explaining Delta measures for authorship attribution". *Digital Scholarship in the Humanities*, 32, pp. ii4-16.
- Fernández López, María (2015). *Obra poética completa*. Martina Vinatea Recoba (ed.). New York: IDEA.
- Fox, Gwyn (2008). *Subtle Subversions: Reading Golden Age Sonnets by Iberian Women*. Washington: Catholic University of America Press.
- Fradejas Rueda, José Manuel (2016). "El análisis estilométrico aplicado a la literatura española: las novelas policíacas históricas". *Caracteres. Estudios culturales y críticos de la esfera digital*, 5. 2, pp. 196-246, <http://revistacaracteres.net/revista/vol5n2noviembre2016/analisis-estilometrico/> [16/01/2025].
- García-Reidy, Alejandro (2019). "Deconstructing the Authorship of *Siempre ayuda la verdad*: A Play by Lope de Vega?" *Neophilologus*, 103, 4, pp. 493-510.
- Gorría, Ana (ed.) (2018). *Antología de poetisas españolas: de la generación del 27 al siglo XV*. Barcelona: Alba.
- Grün, Bettina, y Hornik, Kurt (2011). "Topicmodels: An R Package for Fitting Topic Models". *Journal of Statistical Software*, 40. 1, pp. 1-30. DOI: <https://doi.org/10.18637/jss.v040.i13>.

- Hernández-Lorenzo, Laura (2019). “Poesía áurea, estilometría y fiabilidad: Métodos supervisados de atribución de autoría atendiendo al tamaño de las muestras”. *Caracteres. Estudios culturales y críticos de la esfera digital*, 8. 1, pp. 189-228, <http://revistacaracteres.net/wp-content/uploads/2019/06/Caracteresvol8n1mayo2019-estilometria.pdf> [16/01/2025].
- Hernández-Lorenzo, Laura (2022). “Stylistic Change in Early Modern Spanish Poetry Through Network Analysis (with an Especial Focus on Fernando de Herrera’s Role)”. *Neophilologus*, 106, pp. 397-417. DOI: <https://doi.org/10.1007/s11061-021-09717-2>.
- Hernández-Lorenzo, Laura, y Byszuk, Joanna (2023). “Challenging Stylometry: the authorship of the baroque play *La Segunda Celestina*”. *Digital Scholarship in the Humanities*, 38. 2, pp. 544-58. DOI: <https://doi.org/10.1093/llc/fqac063>.
- Hoover, David L. (2013). “Textual Analysis”. En Kenneth M. Price y Ray Siemens (eds.). *Literary Studies in the Digital Age*. Modern Language Association of America, <https://dlsanthology.mla.hcommons.org/textual-analysis/> [16/01/2025].
- Jannidis, Fotis, *et al.*, (eds.). *Digital Humanities. Eine Einführung*. Stuttgart: J.B. Metzler Verlag, 2017.
- Jockers, Matthew L. (2013). *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press.
- Jockers, Matthew L., y Mimno, David (2013). “Significant Themes in 19th-Century Literature”. *Poetics*, 41. 6, pp. 750-69.
- Jockers, Matthew L., y Underwood, Ted (2016). “Text-Mining the Humanities”. En Susan Schreibman *et al.*, *A New Companion to Digital Humanities*. Wiley-Blackwell, pp. 291-306.
- Koolen, Corina W. (2018). *Reading beyond the Female. The Relationship between Perception of Author Gender and Literary Quality*. Amsterdam: Universiteit van Amsterdam.

- Koppel, Moshe, *et al.* (2022) “Automatically Categorizing Written Texts by Author Gender”. *Literary and Linguistic Computing*, 17, 4, pp. 401-12. DOI: <https://doi.org/10.1093/lc/17.4.401>.
- Kroll, Simon, y Sanz-Lázaro, Fernando (2022). “Romances teatrales entre Mira de Amescua, Calderón y Lope: Ritmo, asonancia y cuestiones de autoría”. *Revista de Humanidades Digitales*, 7, pp. 1-18. DOI: <https://doi.org/10.5944/rhd.vol.7.2022.31620>.
- Lou, Andrés, *et al.* (2015) “Multilabel Subject-Based Classification of Poetry”. *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*. Association for the Advancement of Artificial Intelligence, <https://aaai.org/papers/187-flairs-2015-10372/> [16/01/2025].
- Marcos, Mercedes (2022). *Un mar de sed donde me anego: la obra poética de Ana de la Trinidad*. Burgos: Grupo editorial Fonte.
- Martos, María Dolores (2018). “La voz poética”. En Nieves Baranda y Anne J. Cruz (eds.). *Las escritoras españolas de la Edad Moderna: historia y guía para la investigación*. Madrid: UNED, pp. 225-48.
- Mimno, David, *et al.* (2011). “Optimizing Semantic Coherence in Topic Models”. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh: Association for Computational Linguistics, pp. 262-72, <https://aclanthology.org/D11-1024> [16/01/2025].
- Moretti, Franco (2013). *Distant Reading*. London: Verso.
- Navarro-Colorado, Borja (2018). “On Poetic Topic Modeling: Extracting Themes and Motifs From a Corpus of Spanish Poetry”. *Frontiers in Digital Humanities*, 5, <https://www.frontiersin.org/article/10.3389/fdigh.2018.00015> [16/01/2025].
- Navarro-Colorado, Borja, *et al.* (2016). “Metrical Annotation of a Large Corpus of Spanish Sonnets: Representation, Scansion and Evaluation.” En Nicoletta Calzolari *et al.* (eds.), *Proceedings of the*

- 10th edition of the Language Resources and Evaluation Conference, 23-28 May 2016, Portorož (Slovenia)*. Portorož: European Language Resources Association, pp. 4360-64, <https://aclanthology.org/L16-1691/> [16/01/2025].
- Olivares, Julián y Elizabeth Boyle (eds.) (2021). *Tras el espejo la musa escribe: lírica femenina de los Siglos de Oro*. Madrid: Siglo Veintiuno.
- Pennebaker, James (2011). *The Secret Life of Pronouns: What Our Words Say about Us*. New York: Bloomsbury Press.
- Piper, Andrew (2016). “There Will Be Numbers”. *Cultural Analytics*, 1. 1, pp. 1-10. DOI: <https://doi.org/10.22148/16.006>.
- Ramírez de Guzmán, Catalina (2004). *Obras poéticas*. Joaquín de Entrambasaguas (ed.). Brenes: Muñoz Moya Editores Extremeños.
- Savoy, Jacques (2018). “Is Starnone Really the Author behind Ferrante?” *Digital Scholarship in the Humanities*, 33. 4, pp. 902–918. DOI: <https://doi.org/10.1093/lc/fqy016>.
- Schöch, Christof (2017). “Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama”. *Digital Humanities Quarterly*, 11, 2, <http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html> [16/01/2025].
- Sinclair, Stéfan y Geoffrey Rockwell (2016). “Text Analyzing and Visualization: Making the Meaning Count”. En Susan Schreibman, Ray Siemens y John Unsworth (eds.). *A New Companion to Digital Humanities*. Wiley-Blackwell, pp. 274-290.
- Zayas, María de (2014). *La traición en la amistad*. Teresa Ferrer Valls (ed.). Alicante: Biblioteca Virtual Miguel de Cervantes, <https://www.cervantesvirtual.com/obra/la-traicion-en-amistad/> [16/01/2025].
- Zayas, María de (2022). *Novelas amorosas y ejemplares*. Julián Olivares (ed.). Madrid: Cátedra.

Weidman, Sean G. y O'Sullivan, James (2018). "The limits of distinctive words: Re-evaluating literature's gender marker debate". *Digital Scholarship in the Humanities*, 33. 2, pp. 374-390.