

Una primera aproximación a los géneros literarios de la obra cervantina desde la estilometría y el análisis de redes¹

A first approach to literary genre in Cervantes' works
from Stylometry and Network Analysis

Laura Hernández-Lorenzo

(Universidad Nacional de Educación a Distancia)

RESUMEN

Siguiendo la línea de estudios previos que aplican métodos estilométricos para el estudio de cuestiones estilísticas no autoriales, en este trabajo se realiza una primera aproximación a la obra de Miguel de Cervantes. Mediante la combinación de la estilometría y del análisis de redes, se pretende explorar las relaciones entre sus obras y determinar qué papel juegan la periodización de estas, así como los diferentes géneros literarios que abarcan. Los resultados obtenidos, si bien no muestran una señal cronológica clara, ponen de manifiesto que es posible detectar relaciones de género literario de forma automática a través de estas metodologías. Se realizan, además, nuevas aportaciones sobre las características genéricas de obras concretas, que abren novedosas vías de investigación para el futuro.

PALABRAS CLAVE

Estilometría, análisis de redes, Cervantes, estilística computacional, género literario.

ABSTRACT

Following previous studies applying Stylometry Beyond Authorship Attribution, this paper presents a first approach to the literary works of Miguel de Cervantes. By combining stylometry and network analysis, it aims to explore stylistic connections between his works and to determine the role played by periodisation and different literary genre. The results obtained, although they do not show a clear chronological signal, prove that it is possible to detect literary genre connections automatically through these methodologies. In addition, new contributions are made on the genre characteristics of specific works, which open up new avenues of research for the future.

KEYWORDS

Stylometry, Network Analysis, Cervantes, Computational Stylistics, Literary genre.

1. La publicación es parte de la ayuda FJC2021-047998-I, financiada por MCIN/AEI/10.13039/501100011033 y por la Unión Europea «NextGenerationEU»/PRTR.

Recibido: 13/12/2022

Aceptado: 3/3/2023

1. Introducción

Desde los orígenes de la estilometría han predominado sus aplicaciones a problemas de atribución de autoría. Sin embargo, en los últimos años están aumentando los trabajos que exploran mediante métodos estilométricos otras cuestiones estilísticas, como las corrientes literarias o artísticas (Burrows, 2003), la influencia del sexo del autor (Rybicki, 2016) o el género literario (Jannidis y Lauer, 2014). Los buenos resultados obtenidos muestran el potencial de los métodos y medidas estilométricas para el análisis de corpus textuales desde estas perspectivas y suponen un aliciente para el estudio del estilo literario en general desde el punto de vista estilométrico.

En esta línea, se presenta un análisis mediante técnicas estilométricas y de análisis de redes de las obras completas de un único autor, Miguel de Cervantes (1547-1616). El objetivo final es explorar las relaciones entre sus obras, determinar qué papel juegan la periodización de estas, así como los diferentes géneros literarios que abarcan, y en última instancia, descubrir qué puede aportar esta perspectiva de estudio al conocimiento de la obra cervantina.

La estructura de este artículo es la siguiente: después de esta introducción, se presenta la estilometría como una de las disciplinas más destacadas de las Humanidades Digitales, y se esbozan sus posibles aplicaciones a la obra de Cervantes. Seguidamente, se describe el corpus utilizado y su proceso de preparación, se exponen la metodología, herramientas y parámetros utilizados y los resultados obtenidos. Por último, se presentan las conclusiones de este estudio.

2. La estilometría más allá de la autoría y sus posibles aplicaciones a la obra de Cervantes

La estilometría es una de las disciplinas más destacadas y longevas de las Humanidades Digitales, con orígenes que se remontan al siglo XIX. Habitualmente suele incluirse dentro de los enfoques cuantitativos (Hoover, 2008; Jannidis et al., 2017), y en el último *Companion to Digital Humanities* aparece englobada dentro de la línea de investigación de minería de textos (Jockers y Underwood, 2016). La estilometría y sus métodos se fundamentan en que es posible realizar una aproximación a las características estilísticas de un texto a través de la aplicación de medidas estadísticas a determinados rasgos lingüísticos (Hernández-Lorenzo, 2022a). Desde el exitoso estudio de los *Federalist Papers*, el rasgo que suele analizarse son las palabras más frecuentes, que incluyen las palabras función de un texto, consideradas como un eficaz marcador autorial por su alta frecuencia, su mayor independencia temática y por escapar al control consciente del autor (Stammatos, 2009; Kestemont, 2014). En cuanto a las medidas estadísticas utilizadas, los estudios de evaluación han puesto de manifiesto que la más eficiente a nivel general es la medida *Cosine Delta* (Evert et al., 2015; Evert et al., 2017).

Aunque se ha empleado mayoritariamente para identificar al autor de un texto, la «estilometría más allá de la autoría» (*Stylometry Beyond Authorship Attribution*) tiene un importante precedente en los estudios de Lutosławski sobre la cronología de las obras de Platón (Lutosławski, 1897). Posteriormente, John Burrows, el creador de la popular medida *Delta* (Burrows, 2002),

planteó cómo estos métodos podían ser útiles también para avanzar en otras cuestiones de la historia de la literatura (Burrows, 1987), realizando él mismo análisis estilométricos de distintos géneros y épocas literarias a través de la aplicación de Delta (Burrows, 2003, 2004). Siguiendo a Burrows, Jannidis y Lauer exploran estas cuestiones mediante la aplicación de esta misma medida a un corpus de literatura alemana con el objetivo de evaluar si puede detectar similitudes genéricas y de épocas, con alentadores resultados (Jannidis y Lauer, 2014). Del mismo modo, Hoover (2014) extiende estos experimentos a otros narradores de la literatura inglesa, y Rybicki (2016) aplica estas técnicas para detectar rasgos distintivos en la escritura de hombres y de mujeres. En los últimos años, destacan los trabajos llevados a cabo por los investigadores del proyecto CLiGS, con estudios sobre el género literario en el teatro clásico francés o la narrativa en lengua española de finales del XIX y principios del XX a través de herramientas estilométricas (Calvo Tello, 2019; Henny-Krahmer, 2018; Schöch, 2018). Todos estos trabajos coinciden en que, si bien la señal autorial es la predominante, cuando esta se neutraliza en la medida de lo posible en el corpus, es posible apreciar otras relaciones estilísticas y realizar aportaciones de interés sobre obras y autores a través del análisis cuantitativo de estas.

A pesar del potencial que presenta esta línea de investigación, en el caso de la literatura hispánica, si bien contamos con un creciente número de estudios que aplican métodos estilométricos para resolver cuestiones autoriales (Blasco, 2016, 2019; Reißler-Pipka, 2016; De la Rosa y Suárez, 2016; Calvo Tello y Cerezo Soler, 2018; García-Reidy, 2019; Bazzaco, 2021; Hernández-Lorenzo y Byszuk, 2023), resultan escasos los que se centran en el análisis de otras cuestiones estilísticas, como los movimientos literarios (Hernández-Lorenzo, 2022a) o el género textual (Calvo Tello, 2019, 2021; Hernández-Lorenzo, 2022b). Numerosos investigadores han comprobado en sus estudios que hay diferentes factores o señales estilísticas que pueden afectar a los resultados de un análisis estilométrico, siendo las principales la señal autorial, de género literario, de sexo, de perspectiva narrativa o cronológica. Por esta razón, si se quiere analizar un aspecto estilístico concreto es conveniente neutralizar en la medida de lo posible el resto de las señales en la preparación del corpus. Así, en sus estudios, Jannidis y Lauer (2014) y Calvo Tello (2019) aplican las técnicas estilométricas sobre el corpus de un único autor por análisis con el objetivo de neutralizar la señal autorial y poder explorar el resto de señales estilísticas. En el ámbito de la literatura hispánica, Calvo Tello (2019) ya aplicó este enfoque de forma exitosa al análisis de un corpus de novelas de Valle-Inclán, mientras que la autora de este trabajo analizó la obra completa de Bécquer con novedosos resultados (Hernández-Lorenzo, 2022). Esto supone un aliciente para extender la misma metodología de análisis a la obra de otros autores que cultivaron varios géneros textuales, empezando por los más consagrados, como Cervantes.

Además de ser el autor más emblemático de la literatura española, Miguel de Cervantes ha sido considerado por la crítica como un escritor que desborda los moldes establecidos creando su propia teoría literaria (Riley, 1998). En este sentido, se ha señalado cómo subvierte y transforma los géneros literarios de su época (Close, 1993; Blasco, 2005; Maestro, 2016).² Para Mariscal, como escritor, Cervantes «no se contentaba simplemente con seguir modelos, sino que era capaz de crear géneros nuevos» (Mariscal, 2001: 913). Por su parte, Close señala que uno de los rasgos más modernos del arte cervantino es «fundarse en un diálogo crítico y creador entre géneros literarios» (Close, 1993: 103), y Riley (2001) destaca cómo la conciencia genérica de Cervantes

2. Maestro apunta que, lejos de reproducir los paradigmas de las formas narrativas de su tiempo, Cervantes «subvierte, con frecuencia muy provocativamente, los procedimientos poéticos y canónicos que, en su época histórica, serían en rigor aceptables» (Maestro, 2016: 35).

forma parte de su elevada conciencia crítica. Además del crisol genérico que constituye el *Quijote*, entre otras cuestiones, la crítica ha señalado la presencia de numerosos rasgos teatrales en las *Novelas ejemplares*, donde estudiosos como Close apuntan a «la conversión de la prosa narrativa en teatro»³ (Close, 1993: 99).

En definitiva, este carácter experimental de la obra cervantina, junto a la variedad de géneros cultivados, su particular concepción de la literatura (Canavaggio, 1998; Close, 1998), y su importancia y repercusión, así como su singularidad, refuerzan los atractivos de la aplicación de la estilometría más allá de la autoría para explorar los límites genéricos en su obra completa.

3. Preparación del corpus

Para llevar a cabo este estudio, la obra completa de Cervantes ha sido obtenida del portal dedicado a este escritor en la Biblioteca Virtual Miguel de Cervantes (Bia y Pedreño, 2001).⁴ Todos los textos han sido recopilados en formato de texto plano, para que puedan ser procesados en los análisis posteriores, sin que las etiquetas HTML interfieran en los resultados.

Debido a la breve extensión de las composiciones líricas en verso, estas se mantienen en un único conjunto textual, que abarca, por tanto, distintas fechas. Del mismo modo, las *Novelas ejemplares* se mantienen separadas, ya que son obras independientes. El corpus resultante contiene, pues, 36 obras de diversos géneros y temas: entre otras, la novela pastoril de *La Galatea* (1585), piezas teatrales posiblemente escritas entre 1581 y 1615 (publicadas este último año), obras en verso como *El Viaje del Parnaso* (1614), los dos *Quijotes* (el de 1605 y el 1615), las *Novelas ejemplares*, publicadas en 1613 y que siguen el modelo italiano de la *novella*, y el *Persiles*, su novela bizantina, publicada de forma póstuma en 1617. La lista completa de obras con los géneros, subgéneros a los que pertenecen y su extensión puede consultarse en la Tabla 1.

Título	Clasificación aceptada por la crítica		Palabras
	Género	Subgénero	
<i>El amante liberal</i>	Narrativo	Novela ejemplar	18561
<i>El casamiento engañoso</i>	Narrativo	Novela ejemplar	5163
<i>El celoso extremeño</i>	Narrativo	Novela ejemplar	13699
<i>El coloquio de los perros</i>	Narrativo	Novela ejemplar	23050
<i>El gallardo español</i>	Dramático	Comedia	18362

3. Piqueras también señala similitudes estructurales entre las *Novelas ejemplares* y las partes de comedias, y concluye: «parece lógico pensar que Cervantes, viejo dramaturgo, se valiera de una disposición propia de los libros de comedias para el formato externo de su colección, de la misma forma que había usado otros rasgos teatrales en la redacción de sus novelas, y más teniendo en cuenta que las partes publicadas hasta entonces estaban gozando de un éxito de público considerable. De esta forma, las *Novelas ejemplares* pueden llegar a considerarse como «comedias en prosa» también en este sentido» (Piqueras Flores, 2016: 292).

4. Se han utilizado las ediciones digitales en formato HTML preparadas por Florencio Sevilla Arroyo, que ofrecen una serie de ventajas para el análisis computacional, siendo una de las principales que siguen las mismas normas ortográficas y criterios de edición, garantizando así que estos no interfieran en los resultados. Asimismo, para este primer análisis, se recogen únicamente las obras de autoría segura de Cervantes. En futuros trabajos, se plantea el interés de añadir las obras de autoría dudosa, como *La conquista de Jerusalén* (Cerezo Soler y Calvo Tello, 2019), para ver las conexiones de género literario que estas aportan.

<i>El juez de los divorcios</i>	Dramático	Comedia	3081
<i>El laberinto de amor</i>	Dramático	Comedia	18796
<i>El licenciado Vidriera</i>	Narrativo	Novela ejemplar	9482
<i>El retablo de las maravillas</i>	Dramático	Entremés	3451
<i>El rufián dichoso</i>	Dramático	Comedia	16547
<i>El rufián viudo</i>	Dramático	Entremés	2909
<i>El Trato de Argel</i>	Dramático	Comedia	15552
<i>El viejo celoso</i>	Dramático	Entremés	3980
<i>El vizcaíno fingido</i>	Dramático	Entremés	4557
<i>La Galatea</i>	Narrativo	Novela pastoril	117969
<i>La casa de los celos</i>	Dramático	Comedia	16630
<i>La cueva de Salamanca</i>	Dramático	Entremés	3721
<i>La elección de los alcaldes de Daganzo</i>	Dramático	Entremés	2541
<i>La entretenida</i>	Dramático	Comedia	16972
<i>La española inglesa</i>	Narrativo	Novela ejemplar	16041
<i>La fuerza de la sangre</i>	Narrativo	Novela ejemplar	7696
<i>La gitanilla</i>	Narrativo	Novela ejemplar	23472
<i>La guarda cuidadosa</i>	Dramático	Entremés	3878
<i>La gran sultana</i>	Dramático	Comedia	16657
<i>La ilustre fregona</i>	Narrativo	Novela ejemplar	20016
<i>La Numancia</i>	Dramático	Tragedia renacentista	16042
<i>Las dos doncellas</i>	Narrativo	Novela ejemplar	14971
<i>La señora Cornelia</i>	Narrativo	Novela ejemplar	14352
<i>Los baños de Argel</i>	Dramático	Comedia	18265
<i>Pedro de Urdemalas</i>	Dramático	Comedia	16970
<i>El Persiles</i>	Narrativo	Novela bizantina	142710
<i>Poesía</i>	Poesía	Poesía lírica	10404
<i>Primera parte del Quijote</i>	Narrativo	Novela	181178
<i>Segunda parte del Quijote</i>	Narrativo	Novela	191759
<i>Rinconete y Cortadillo</i>	Narrativo	Novela ejemplar	13696
<i>El viaje del Parnaso</i>	Poesía	Poema narrativo	24802

Tabla 1. Corpus de obras de Miguel de Cervantes utilizado

4. Metodología

Este estudio utiliza la metodología de la estilometría más allá de la autoría, siguiendo los trabajos citados en el estado de la cuestión (sección 2). Se combinan métodos clásicos de esta disciplina, como los análisis de grupos y los árboles de consenso (Eder, 2013), con los análisis de redes estilométricas siguiendo la implementación de Eder (2017).

El primero de los métodos empleados, el análisis de grupos, es una técnica estadística que trata de encontrar los textos estilísticamente más similares mediante la creación de una jerarquía de grupos que parte de los textos más cercanos y semejantes entre ellos, hasta los más lejanos y diferentes entre sí (Eder, 2017: 51). En el segundo método varios análisis de grupos realizados con diversos parámetros son combinados en un único árbol de consenso, que contiene las relaciones más estables entre los textos, es decir, aquellas que permanecen a lo largo de diferentes parámetros (Eder, 2013). El tercer y último método se caracteriza por el hecho de que permite visualizar y analizar un gran número de relaciones estilísticas entre textos a través del análisis de redes. Como explica Eder (2017), puesto que los análisis de redes estilométricas mantienen un mayor número de conexiones entre textos (no solo las más consistentes, sino también otras más débiles), resultan de especial utilidad para explorar cuestiones estilísticas diferentes a la detección autorial.

El análisis de grupos y el árbol de consenso han sido realizados mediante el paquete *stylo* en lenguaje R (Eder et al., 2016), y se ha utilizado el programa Gephi (Bastian et al., 2009) para convertir el *output* del árbol de consenso en un grafo. En cuanto a los parámetros del análisis, se ha utilizado Cosine Delta (Evert et al., 2017) por ser la medida de distancia que generalmente obtiene mejores resultados, aplicada sobre las 100 palabras más frecuentes en el análisis de grupos y sobre las 100 a 1000 palabras más frecuentes en el árbol de consenso. Este rango ha sido elegido siguiendo los parámetros habitualmente utilizados y con el objetivo de restringir el análisis en la medida de lo posible a las palabras función o agramaticales, de forma que se produzcan menos interferencias temáticas en los resultados.

5. Resultados

5.1. Análisis de grupos

Para comenzar, se muestra el resultado del análisis de grupos o *cluster analysis* en la Figura 1. Como puede observarse, las obras de Cervantes quedan divididas fundamentalmente en dos grandes grupos: un primero que incluye sus textos dramáticos y poéticos, y un segundo con sus textos narrativos. De este modo, en el segundo grupo encontramos las *Novelas ejemplares*, las dos partes del *Quijote*, *La Galatea* y el *Persiles*; en cambio, el primer grupo contiene todas las comedias y entremeses compuestos por Cervantes, junto a su poesía lírica y el *Viaje del Parnaso*, formando estas dos últimas un pequeño subgrupo. La fuerte división entre ambos grupos queda reflejada en la amplitud numérica de la escala del gráfico (0-10).

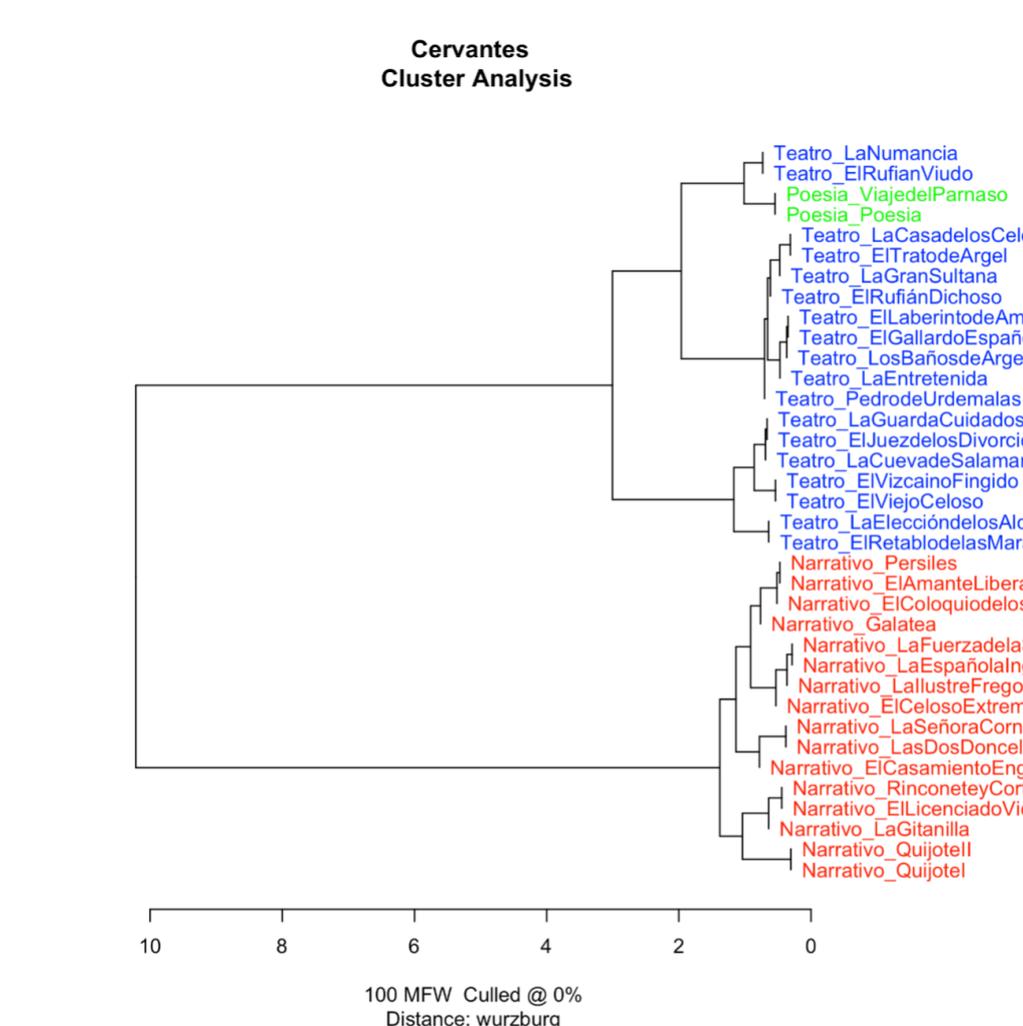


Figura 1. Análisis de grupos de la obra completa de Cervantes aplicando Cosine Delta a las 100 palabras más frecuentes

5.2. Árbol de consenso

Seguidamente, la Figura 2 presenta el árbol de consenso aplicando Cosine Delta sobre las 100 a 1000 palabras más frecuentes en la obra cervantina. Como es sabido, esta técnica recoge las conexiones más estables entre textos a largo de una serie de iteraciones. Como resultado, se observan menos relaciones entre textos y algunos de ellos cuelgan directamente de la raíz, como es el caso de varias *Novelas Ejemplares* (*El amante liberal*, *El casamiento engañoso*, *El celoso extremeño*, *La gitanilla*, *La ilustre fregona* y *Rinconete y Cortadillo*) y del *Persiles*. En este caso, se confirma que el dendograma es más interesante que el árbol de consenso para explorar cuestiones estilísticas no relacionadas con la detección de autoría, tal y como apuntan algunos estudios previos (Jannidis y Lauer, 2014; Hernández-Lorenzo, 2022b). A pesar de ello, se mantiene la estrecha relación entre la poesía lírica y el *Viaje del Parnaso*, y encontramos tres ramas con las obras teatrales de Cervantes, una de las cuales incluye todos sus entremeses, otra las comedias publicadas en 1615 y la tercera, las dos tragedias de su primera etapa (*La Numancia* y *El Trato de Argel*). Del mismo modo,

los dos *Quijotes* aparecen juntos, así como algunas de las *Novelas ejemplares*. Resulta curioso que *La Galatea* aparezca junto a la poesía, seguramente debido a la abundancia de textos poéticos en esta, donde la poesía se convierte en asunto central (Montero, 2016).⁵

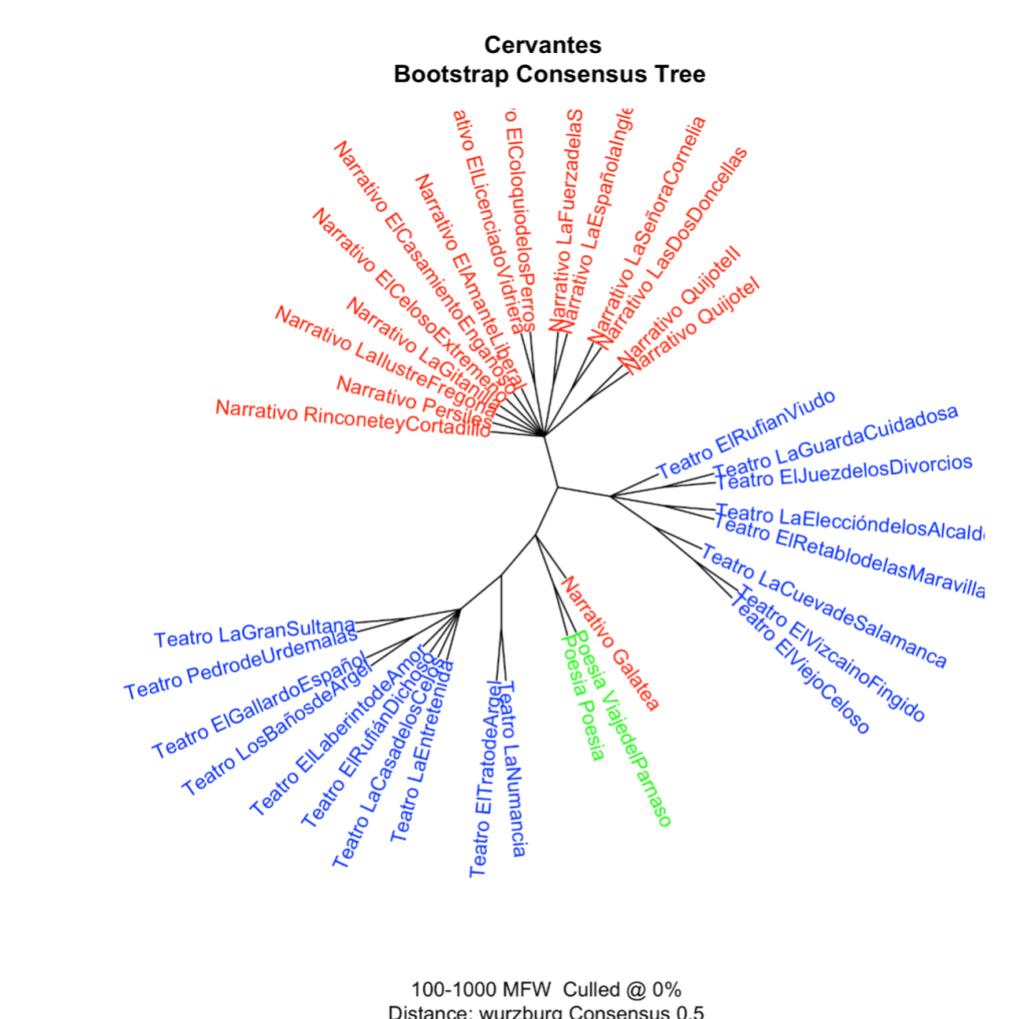


Figura 2. Árbol de consenso de la obra completa de Cervantes aplicando Cosine Delta sobre las 100 a 1000 palabras más frecuentes

5.3. Red estilométrica

Por último, la Figura 3 muestra la red estilométrica obtenida con el *output* para Gephi generado por *stylo* del árbol de consenso anterior. Como puede apreciarse, las obras de Cervantes forman una red alargada, con el extremo izquierdo dominado por las obras teatrales y el derecho por las narrativas. El factor determinante en la formación de la red no parece ser la cronología, puesto que una obra temprana como *La Galatea* aparece junto a la novela bizantina póstuma de Cervantes, el

5. Montero (2016) señala cómo la poesía desempeña un papel fundamental y no solo porque, del mismo modo que en otras novelas pastoriles, en *La Galatea* encontramos versos recitados por distintos personajes, contabilizándose más de 5000, sino también por la destacada presencia en la obra de poetas españoles del momento, ya sea disfrazados en un contexto pastoril o mencionados con sus nombres reales.

Persiles. La aplicación del algoritmo de modularidad ha permitido distinguir cuáles son las principales comunidades de la red. Con una resolución de 1, pueden distinguirse cinco comunidades: las dos de color verde claro y oscuro contienen las obras teatrales, mientras que la rosa y la naranja están constituidas por los textos narrativos. La quinta comunidad, en color celeste, se localiza en el centro de la red, y en ella se encuentra la poesía lírica y narrativa de Cervantes junto a dos de sus obras teatrales (*La Numancia* y *El trato de Argel*). Es interesante cómo estas dos últimas, a pesar de encontrarse cerca de las comunidades teatrales, parecen estar más cercanas a la poesía que el resto. Tanto estas dos comedias más tempranas como las publicadas en 1615 están escritas en verso, por lo que el resultado obtenido parece indicar un mayor aliento poético de estas dos piezas.

Además, si se presta atención a las dos comunidades teatrales, se advierte que la comunidad de color verde claro agrupa los entremeses, mientras que la de color verde oscuro contiene las comedias. En este sentido, la única excepción es *El rufián viudo*, que aparece en esta última comunidad a pesar de ser un entremés. En cuanto a las comunidades narrativas, cabe destacar cómo la de color rosa parece constituida por las obras más cómicas y grotescas, como la novela ejemplar picaresca de *Rinconete y Cortadillo* o los dos *Quijotes*, mientras que la de color naranja muestra obras con un tono más elevado o más idealizadas como la novela pastoril de *La Galatea*, la novela bizantina del *Persiles* o la novela ejemplar de *La española inglesa*, *Las dos doncellas*, *El amante liberal* o *La señora Cornelia*.⁶

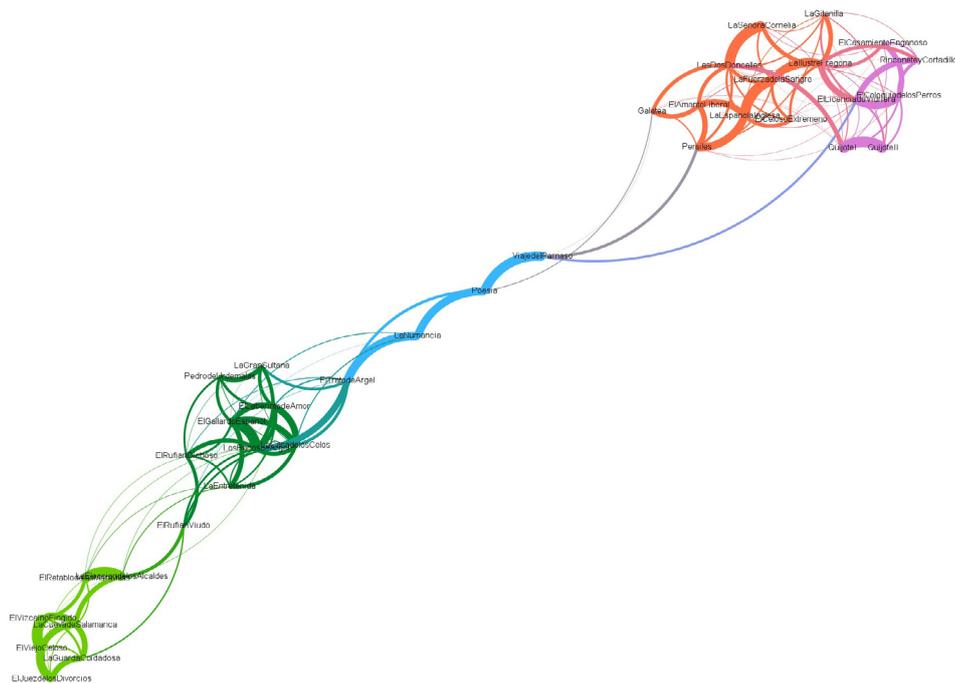


Figura 3. Análisis de redes de la obra completa de Miguel de Cervantes coloreado por Modularidad, con una resolución de 1

6. Montero señala cómo estas obras pertenecen «a la nebulosa de lo que se suele llamar narrativa idealista» (2016: 106).

5.4. *Resumen de resultados*

Para una mejor comprensión de los resultados obtenidos, estos se han incluido de forma esquemática en la Tabla 2, junto a las clasificaciones de las obras según su género y subgénero.

Título	Clasificación aceptada por la crítica		Resultados estilométricos	
	Género	Subgénero	Método de análisis de grupos	Método de red estilométrica
<i>El amante liberal</i>	Narrativo	Novela ejemplar	Grupo narrativo	1ª comunidad narrativa
<i>El casamiento engañoso</i>	Narrativo	Novela ejemplar	Grupo narrativo	2ª comunidad narrativa
<i>El celoso extremeño</i>	Narrativo	Novela ejemplar	Grupo narrativo	1ª comunidad narrativa
<i>El coloquio de los perros</i>	Narrativo	Novela ejemplar	Grupo narrativo	2ª comunidad narrativa
<i>El gallardo español</i>	Dramático	Comedia	Grupo dramático	1ª comunidad dramática: comedias
<i>El juez de los divorcios</i>	Dramático	Comedia	Grupo dramático	2ª comunidad dramática: entremeses
<i>El laberinto de amor</i>	Dramático	Comedia	Grupo dramático	1ª comunidad dramática: comedias
<i>El licenciado Vidriera</i>	Narrativo	Novela ejemplar	Grupo narrativo	2ª comunidad narrativa
<i>El retablo de las maravillas</i>	Dramático	Entremés	Grupo dramático	2ª comunidad dramática: entremeses
<i>El rufián dichoso</i>	Dramático	Comedia	Grupo dramático	1ª comunidad dramática: comedias
<i>El rufián viudo</i>	Dramático	Entremés	Grupo dramático	1ª comunidad dramática: comedias
<i>El Trato de Argel</i>	Dramático	Comedia	Grupo dramático	Comunidad lírica
<i>El viejo celoso</i>	Dramático	Entremés	Grupo dramático	2ª comunidad dramática: entremeses

<i>El vizcaíno fingido</i>	Dramático	Entremés	Grupo dramático	2ª comunidad dramática: entremeses
<i>La Galatea</i>	Narrativo	Novela pastoril	Grupo narrativo	1ª comunidad narrativa
<i>La casa de los celos</i>	Dramático	Comedia	Grupo dramático	1ª comunidad dramática: comedias
<i>La cueva de Salamanca</i>	Dramático	Entremés	Grupo dramático	2ª comunidad dramática: entremeses
<i>La elección de los alcaldes de Daganzo</i>	Dramático	Entremés	Grupo dramático	2ª comunidad dramática: entremeses
<i>La entretenida</i>	Dramático	Comedia	Grupo dramático	1ª comunidad dramática: comedias
<i>La española inglesa</i>	Narrativo	Novela ejemplar	Grupo narrativo	1ª comunidad narrativa
<i>La fuerza de la sangre</i>	Narrativo	Novela ejemplar	Grupo narrativo	1ª comunidad narrativa
<i>La gitanilla</i>	Narrativo	Novela ejemplar	Grupo narrativo	1ª comunidad narrativa
<i>La guarda cuidadosa</i>	Dramático	Entremés	Grupo dramático	2ª comunidad dramática: entremeses
<i>La gran sultana</i>	Dramático	Comedia	Grupo dramático	1ª comunidad dramática: comedias
<i>La ilustre fregona</i>	Narrativo	Novela ejemplar	Grupo narrativo	1ª comunidad narrativa
<i>La Numancia</i>	Dramático	Tragedia renacentista	Grupo dramático	Comunidad lírica
<i>Las dos doncellas</i>	Narrativo	Novela ejemplar	Grupo narrativo	1ª comunidad narrativa
<i>La señora Cornelia</i>	Narrativo	Novela ejemplar	Grupo narrativo	1ª comunidad narrativa
<i>Los baños de Argel</i>	Dramático	Comedia	Grupo dramático	1ª comunidad dramática: comedias

<i>Pedro de Urdemalas</i>	Dramático	Comedia	Grupo dramático	1ª comunidad dramática: comedias
<i>El Persiles</i>	Narrativo	Novela bizantina	Grupo narrativo	1ª comunidad narrativa
Poesía	Poesía	Poesía lírica	Grupo dramático. Subgrupo lírico	Comunidad lírica
<i>Primera parte del Quijote</i>	Narrativo	Novela	Grupo narrativo	2ª comunidad narrativa
<i>Segunda parte del Quijote</i>	Narrativo	Novela	Grupo narrativo	2ª comunidad narrativa
<i>Rinconete y Cortadillo</i>	Narrativo	Novela ejemplar	Grupo narrativo	2ª comunidad narrativa
<i>El viaje del Parnaso</i>	Poesía	Poema narrativo	Grupo dramático. Subgrupo lírico	Comunidad lírica

Tabla 2. Clasificaciones de las obras cervantinas según su género, subgénero y los resultados obtenidos en los análisis estilométricos (análisis de grupos y red estilométrica)

5.5. *Discusión e interpretación*

Como puede apreciarse en la tabla anterior, los resultados estilométricos coinciden en la mayoría de los casos con las clasificaciones previas, pero se detectan algunos casos que merecen un comentario más detenido.

En cuanto a las dos comunidades obtenidas en la red estilométrica para los textos narrativos, estas pueden relacionarse con la clasificación de Riley (2001) entre novelas y romances. Este estudioso defiende la coexistencia en la prosa de ficción de Cervantes de dos tipos de escritura: la ficción realista en prosa (las novelas) y la romántica e idealista (los romances). En lugar de producirse una evolución de una forma hacia la otra, ambas se encontrarían en la obra cervantina, bien separadas o mezcladas en diferentes grados. De acuerdo con esta interpretación, Riley realiza una clasificación de las obras narrativas donde predomina uno de estos dos tipos o donde se encuentran mezclados, que puede consultarse en la Tabla 3 junto a los resultados obtenidos en la red estilométrica.

Obra narrativa	Clasificación siguiendo a Riley	Resultados de la red estilométrica
<i>La Galatea</i>	Romance	1ª comunidad narrativa: romances
<i>Persiles</i>	Romance	1ª comunidad narrativa: romances
<i>El amante liberal</i>	Romance	1ª comunidad narrativa: romances

<i>La fuerza de la sangre</i>	Romance	1ª comunidad narrativa: romances
<i>La española inglesa</i>	Romance	1ª comunidad narrativa: romances
<i>Las dos doncellas</i>	Romance	1ª comunidad narrativa: romances
<i>La señora Cornelia</i>	Romance	1ª comunidad narrativa: romances
<i>Primera parte del Quijote</i>	Novela	2ª comunidad narrativa: novelas
<i>Segunda parte del Quijote</i>	Novela	2ª comunidad narrativa: novelas
<i>Rinconete y Cortadillo</i>	Novela	2ª comunidad narrativa: novelas
<i>El licenciado Vidriera</i>	Novela	2ª comunidad narrativa: novelas
<i>El celoso extremeño</i>	Novela	1ª comunidad narrativa: romances
<i>El casamiento engañoso</i>	Novela	2ª comunidad narrativa: novelas
<i>El coloquio de los perros</i>	Novela	2ª comunidad narrativa: novelas
<i>La gitanilla</i>	Mixta	1ª comunidad narrativa: romances
<i>La ilustre fregona</i>	Mixta	1ª comunidad narrativa: romances

Tabla 3. Clasificaciones de las obras narrativas de Cervantes siguiendo a Riley (1998) y los resultados obtenidos en la red estilométrica

Como puede comprobarse, la clasificación de Riley (1998) resulta de utilidad para interpretar los resultados obtenidos en la red estilométrica, pues en la mayoría de los casos, la primera comunidad narrativa obtenida coincide con las obras clasificadas por este estudioso como romances, mientras que las clasificadas como novelas corresponden con la segunda comunidad narrativa.⁷ Solo hay tres excepciones: las novelas ejemplares de *El celoso extremeño*, *La gitanilla* y *La ilustre fregona*. Estas tres aparecen en la comunidad de romances, a pesar de que, de acuerdo con Riley, la primera debería clasificarse como novela, y las otras dos constituyen formas mixtas. Si se observan las conexiones que presentan estos textos dentro de la red, *La gitanilla* y *La ilustre fregona* poseen un número similar de conexiones con las obras de la comunidad romance y de la comunidad narrativa (lo cual puede relacionarse con el carácter mixto que les adjudica Riley)⁸, mientras que *El celoso extremeño* posee conexiones mayoritariamente con obras de la comunidad

7. Blasco (2005), sin embargo, matiza que no es el binomio novela / romance el que para Cervantes centra el problema, sino el binomio literatura y vida, entre la narración de lo que podría suceder y lo que había sucedido.

8. Más concretamente, *La gitanilla* aparece conectada con 5 obras de la comunidad romance (*La señora Cornelia*, *Las dos doncellas*, *La ilustre fregona*, *La española inglesa* y *El Persiles*) y 4 obras de la comunidad novelística (*Rinconete y Cortadillo*, *El coloquio de los perros*, *El licenciado Vidriera* y el *Quijote II*), mientras que *La ilustre fregona* posee conexiones con 6 obras de la comunidad romance (*La gitanilla*, *La señora Cornelia*, *Las dos doncellas*, *La fuerza de la sangre*, *La española inglesa* y *El celoso extremeño*) y 5 obras de la comunidad novelística (*El casamiento engañoso*, *Rinconete y Cortadillo*, *El licenciado Vidriera*, y los dos *Quijotes*).

romance, aunque también con dos concretas de la comunidad novelística (*Rinconete y Cortadillo* y *El licenciado Vidriera*).

En otro orden de cosas, resulta llamativa la inclusión de *La Numancia* y el *Trato de Argel* en la comunidad lírica, junto a la poesía de Cervantes. Como es bien sabido, ambas obras se diferencian del resto de su teatro en que pertenecen a la primera etapa del escritor, entre los años 1582 y 1587, un momento de transición anterior a la irrupción del teatro de Lope de Vega (Riquer, 2003). En el prólogo a las *Ocho comedias y ocho entremeses*, el propio Cervantes afirma que se estrenaron con moderado éxito, y destaca los cambios realizados en la estructura de las mismas y el uso del procedimiento de la alegoría (Cervantes, 1615). Armas (2012) profundiza en la importancia del uso de este recurso tanto en *La Numancia* como en *El Trato de Argel*, mientras que otros estudiosos recalcan el carácter metateatral y metapoético de la primera (Gilabert, 2016; Valencia, 2016), y Martos (1990) defiende el interés de la versificación y las rimas y la función esencial que juegan en la obra. Estas particularidades podrían explicar que se separen en la red de las comunidades dramáticas y aparezcan junto a los textos poéticos.

Por último, respecto a las dos comunidades dramáticas, es digna de mención la aparición de *El rufián viudo* en la comunidad de comedias (si bien en el árbol de consenso sí se agrupaba con los entremeses). Esto puede deberse a varias razones: las más probables parecen ser el hecho de que sea uno de los pocos entremeses escritos en verso, y la existencia de interferencias temáticas con la comedia de *El rufián dichoso*. Además, algunos estudiosos destacan su originalidad y señalan que, «pese a ser una obra «menor», participa de la dimensión metaliteraria o autorreflexiva de otros textos cervantinos más ambiciosos» (Checa, 1986: 268).

6. Conclusiones y futuras líneas de investigación

El análisis estilométrico sobre la obra de Cervantes ha arrojado algunos resultados de gran interés:

En primer lugar, los métodos estilométricos han conseguido detectar conexiones de género literario en el corpus, en línea con los resultados obtenidos en los estudios anteriores con otras obras, autores y literaturas. La cronología de las obras, en cambio, no parece ser un factor determinante, ya que obras tempranas como *La Galatea* aparecen junto a otras muy posteriores, como *El Persiles*.

En segundo lugar, la señal de género literario parece ser la más fuerte en la obra cervantina, pues ha dominado tanto la formación del análisis de grupos como de la red estilométrica, y se han obtenido resultados mayoritariamente en línea con las clasificaciones genéricas aceptadas por la crítica. Aunque se aprecia una cierta influencia de la temática en la formación de las comunidades de textos narrativos detectadas mediante el cálculo de la modularidad, y que parecen separar las obras de corte realista de las más idealistas, estas pueden explicarse por la tipología romance / novela (Riley, 2001). Sin embargo, como se ha visto, tres novelas ejemplares (*La gitanilla*, *La ilustre fregona* y *El celoso extremeño*) no encajan en el lugar esperado, lo cual abre el camino a nuevas investigaciones.

En tercer lugar, otra aportación interesante en términos de géneros literarios y que también necesitará un estudio más extenso en el futuro es la estrecha relación entre la poesía cervantina y dos de sus obras teatrales, la tragedia renacentista *La Numancia* y la comedia *El trato de Argel*, que

sugiere un mayor aliento lírico de estas composiciones. Del mismo modo, los resultados apoyan la particularidad de *El rufián viudo* dentro de los entremeses de Cervantes.

Por último, las similitudes señaladas por algunos críticos entre las *Novelas ejemplares* y las comedias cervantinas no se aprecian en los resultados obtenidos, que muestran una clara separación entre la obra narrativa y la teatral.

Bibliografía

- ARMAS, Frederick A. de (2012), «El saber de Herebo/Proteo: La alegoría en *El Trato de Argel* y *La Numancia*», en *Recreaciones teatrales y alegorías cervantinas*, ed. C. Mata Induráin, Pamplona, EUNSA, pp. 145-160.
- BASTIAN, Mathieu, Sebastien HEYMANN y Mathieu JACOMY (2009), «Gephi: An Open Source Software for Exploring and Manipulating Networks», *Proceedings of the International AAAI Conference on Web and Social Media*, 3, 1, pp. 361-362. < <https://doi.org/10.1609/icwsm.v3i1.13937> > [19/03/2009; consulta 28/06/2023].
- 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 «*Prenga xascú ço qui millor li és de mon dit*». *Creació, recepció i representació de la literatura medieval*, ed. M. Simó, G. Avenozza, A. Contreras, G. Sabaté, y L. Soriano, San Millán de la Cogolla, Cilengua, pp. 149-166.
- BIA, Alejandro y Andrés PEDREÑO (2001), «The Miguel de Cervantes Digital Library: The Hispanic Voice on the Web», *Digital Scholarship in the Humanities*, 16, 1, pp. 161-177.
- BLASCO, Javier (2005), *Cervantes, raro inventor*, Alcalá de Henares, Centro de Estudios Cervantinos.
- _____ (2016), «Avellaneda desde la estilometría», en *Cervantes. Los viajes y los días*, ed. P. Ruiz Pérez, Madrid, Prosa Barroca y SIAL Ediciones, pp. 97-116.
- _____ (2019), «Atribuciones cervantinas desde la estilometría: El entremés de *Los mirones*», en *Cartografía teatral: En homenaje al profesor José Romera Castillo*, ed. G. Laín Corona y R. Santiago Nogales, Madrid, Visor, pp. 151-168).
- BURROWS, John (1987), *Computation into Criticism: A Study of Jane Austen’s Novels and an Experiment in Method*, Oxford, Oxford University Press.
- _____ (2002), «Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship», *Literary and Linguistic Computing*, 17, 3, pp. 267-287.
- _____ (2003), «Questions of Authorship: Attribution and Beyond: A Lecture Delivered on the Occasion of the Roberto Busa Award ACH-ALLC 2001, New York», *Computers and the Humanities*, 37, 1, pp. 5-32.
- _____ (2004), «Textual analysis», en *A Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens y J. Unsworth, Oxford, Blackwell, pp. 323-347.
- CALVO TELLO, José (2019), «Delta Inside Valle-Inclan: Stylometric Classification of Periods and Groups of His Novels», *Romanische Studien. Theorien von Autorschaft und Stil in Bewegung: Stilistik und Stilometrie in der Romania*, 6, pp. 151-164.
- _____ (2021), *The Novel in the Spanish Silver Age. A Digital Analysis of Genre using Machine Learning*, Bielefeld, Bielefeld University Press.

- ____ y Juan CEREZO SOLER (2018), «La conquista de Jerusalén ¿de Cervantes? Análisis estilométrico sobre autoría en el teatro del Siglo de Oro español», *Digital Humanities Quarterly*, 12, 1, pp. 1-10. <<http://digitalhumanities.org:8081/dhq/vol/12/1/000346/000346.html>>
- CANAVAGGIO, Jean (1977), *Cervantès dramaturge: Un théâtre à naître*, Paris, PUF.
- ____ (1998), «Vida y literatura: Cervantes en el *Quijote*», en *Don Quijote de la Mancha*, ed. F. Rico, Madrid, Instituto Cervantes, pp. 41-66.
- CEREZO SOLER, Juan y José CALVO TELLO (2019), «Autoría y estilo. Una atribución cervantina desde las humanidades digitales. El caso de *La conquista de Jerusalén*», *Anales Cervantinos*, 51, pp. 231-250.
- CERVANTES, Miguel de (1615), *Ocho comedias y ocho entremeses nuevos nunca representados*, Madrid, por la viuda de Alonso Martín.
- CHECA, Jorge (1986), «*El rufián viudo* de Cervantes: Estructura, imágenes, parodia, carnavalización», *Hispanic Issue*, 101, 2, pp. 247-269.
- CLOSE, Anthony (1993), «Cervantes frente a los géneros cómicos del siglo XVI», en *Actas del III Coloquio Internacional de la Asociación de Cervantistas*, Alcalá de Henares, Anthropos, pp. 89-104.
- ____ (1998), «Cervantes: Pensamiento, personalidad y cultura», en *Don Quijote de la Mancha*, ed. F. Rico, Madrid, Instituto Cervantes, pp. 67-68.
- EDER, Maciej (2013), «Computational stylistics and Biblical translation: How reliable can a dendrogram be?», en *The Translator and the Computer*, ed. T. Piotrowski y L. Grabowski, Wrocław, Wyd. Wyższej Szkoły Filologicznej, pp. 155-170.
- ____ (2017), «Visualization in stylometry: Cluster analysis using networks», *Digital Scholarship in the Humanities*, 32, 1, pp. 50-64.
- ____, RYBICKI, Jan y Mike KESTEMONT (2016), «Stylometry with R: A Package for Computational Text Analysis», *The R Journal*, 8, 1, pp. 107-121.
- EVERT, Stefan, Thomas PROISL, Thorsten VIT, Christof SCHÖCH, Fotis JANNIDIS y Steffen PIELSTRÖM (2015), «Towards a better understanding of Burrows's Delta in literary authorship attribution», en *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, Denver, Association for Computational Linguistics, pp. 79-88. <<https://aclanthology.org/W15-0709>>
- ____, Thomas PROISL, Fotis JANNIDIS, Isabella REGER, Steffen PIELSTRÖM, Christof SCHÖCH y Thorsten VIT (2017), «Understanding and explaining Delta measures for authorship attribution», *Digital Scholarship in the Humanities*, 32, supl. 2, pp. ii4-ii16.
- GARCÍA-REIDY, Alejandro (2019), «Deconstructing the Authorship of *Siempre ayuda la verdad: A Play by Lope de Vega?*», *Neophilologus*, 103, 4, pp. 493-510.
- GILABERT, Gaston (2016), «La construcción especular de la tragedia cervantina», *Anuario de Estudios Cervantinos. Cervantes y los géneros literarios*, 12, pp. 99-114.
- HENNY-KRAHMER, Ulrike (2018), «Exploration of Sentiments and Genre in Spanish American Novels», *Digital Humanities 2018 Puentes-Bridges*, pp. 399-403. <https://dh2018.adho.org/wp-content/uploads/2018/06/dh2018_abstracts.pdf> [consulta 28/06/2023].
- HERNÁNDEZ-LORENZO, Laura (2022a), «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.

- ____ (2022b), «La prosa de Gustavo Adolfo Bécquer en los límites de la poesía: Análisis estilométrico», *apropos. Perspektiven auf die Romania*, 9, pp. 37-56. <<https://doi.org/10.15460/apropos.9.1875>>
- ____ y Joanna BYSZUK (2023), «Challenging Stylometry: The authorship of the baroque play *La Segunda Celestina*», *Digital Scholarship in the Humanities*, 38, 2, pp. 544-558.
- HOOVER, David L. (2008), «Quantitative Analysis and Literary Studies», en *A Companion to Digital Literary Studies*, ed. R. Siemens y S. Schreibman, Oxford, Blackwell, pp. 517-533.
- ____ (2014), «A Conversation Among Himselves: Change and the Styles of Henry James», en *Digital Literary Studies: Corpus Approaches to Poetry, Prose and Drama. The Corpus, the Computer and the Study of Literature*, D. L. Hoover, J. Culpeper y K. O'Halloran, New York, Routledge Advances in Corpus Linguistics, pp. 90-119.
- JANNIDIS, Fotis, Hubertus KOHLE, Malte REHBEIN (ed.) (2017), *Digital Humanities: Eine Einführung*, Stuttgart, J.B. Metzler Verlag.
- ____ y Gerhard LAUER (2014), «Burrows's Delta and Its Use in German Literary History», en *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*, ed. M. Erlin y L. Tatlock, Rochester, Camden House, pp. 29-54.
- JOCKERS, Matthew L. y Ted UNDERWOOD (2016), «Text-Mining the Humanities», en *A New Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, & J. Unsworth, Oxford, Wiley-Blackwell, pp. 291-306.
- KESTEMONT, Mike (2014), «Function Words in Authorship Attribution From Black Magic to Theory?», en *3rd Workshop on Computational Linguistics for Literature (CLfL 2014), January 2014*, Gothenburg, Association for Computational Linguistics, pp. 59-66. <<http://aclweb.org/anthology//W/W14/W14-09.pdf>>
- LUTOSŁAWSKI, Wincenty (1897), *The origin and growth of Plato's logic: With an account of Plato's style and of the chronology of his writings*, London, Longmans, Green & Co.
- MARISCAL, Beatriz (2001), «Cervantes creador de géneros literarios: El desarrollo de un motivo narrativo», en *Actas del IV Congreso Internacional de la Asociación de Cervantistas: Lepanto, Grecia, 1-8 de octubre de 2000*, ed. Antonio Bernat Vistarini, Palma, Universitat de les Illes Balears, pp. 913-919.
- MONTERO, Juan (2016), «*La Galatea*: Hacia Cervantes», *Revista de Occidente*, 427, pp. 106-120.
- PORTAL MIGUEL DE CERVANTES (Alicante, Biblioteca Virtual Miguel de Cervantes) <https://www.cervantesvirtual.com/portales/miguel_de_cervantes> [consulta 26/09/2023].
- PIQUERAS FLORES, Manuel (2016), «Las Ejemplares: Primera parte de novelas de Cervantes», *Anuario de Estudios Cervantinos. Cervantes y los géneros literarios*, 12, pp. 283-295.
- RILEY, Edward C. (1998), «Cervantes: Teoría literaria», en *Don Quijote de la Mancha*, ed. F. Rico, Madrid, Instituto Cervantes, pp. 129-141.
- ____ (2001), *La rara invención. Estudios sobre Cervantes y su posteridad literaria*, Barcelona, Crítica.
- RIQUER, Martín de (2003), *Para leer a Cervantes*, Barcelona, Acanalado.
- RISSELER-PIPKA, Nanette (2016), «Avellaneda y los problemas de la identificación del autor. Propuestas para una investigación con nuevas herramientas digitales», en *El otro Don Quijote. La continuación de Fernández de Avellaneda y sus efectos*, ed. H. Ehrlicher, Augsburg, Universität Augsburg, pp. 27-51.
- ROSA, Javier de la, SUÁREZ, Juan Luis (2016), «*The Life of Lazarillo de Tormes* and of His Machine Learning Adversities. Non-traditional authorship attribution techniques in the context of the *Lazarillo*», *Lemir*, 20, pp. 373-438.

- RYBICKI, Jan (2016), «Vive la différence: Tracing the (authorial) gender signal by multivariate analysis of word frequencies», *Digital Scholarship in the Humanities*, 31, 4, pp. 746-761.
- SCHÖCH, Christof (2018), «Zeta für die kontrastive Analyse literarischer Texte. Theorie, Implementierung, Fallstudie», en *Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*, ed. T. Bernhart, S. Richter, M. Lepper, M. Willand y A. Albrecht, Berlin, De Gruyter, pp. 77-94.
- STAMATATOS, Efstathios (2009), «A survey of modern authorship attribution methods», *Journal of the American Society for information Science and Technology*, 60, 3, pp. 538-556. <<https://onlinelibrary.wiley.com/doi/epdf/10.1002/asi.21001>>