

# Fernando de Herrera y la autoría de *Versos*

## Un primer acercamiento al drama textual desde la Estilometría

Laura Hernández Lorenzo (Sevilla)

**RESUMEN:** Además de tratarse de uno de los problemas textuales y de autoría más fascinantes de la poesía del Siglo de Oro, la cuestión en torno a la autenticidad de parte de la obra lírica de Fernando de Herrera ha dividido a la crítica en dos grandes grupos, dando lugar a un *impasse* que continúa hasta hoy. En este artículo se pretende arrojar nueva luz aplicando metodologías provenientes de las Humanidades Digitales y la Estilometría, empleando una técnica desarrollada recientemente por Maciej Eder, Rolling Stylometry, para detectar si hay varias manos en las obras poéticas dubitadas de Herrera, además de confirmar o rechazar la existencia de la mano de Francisco de Pacheco, principal argumento de los estudiosos que niegan la autoría herreriana.

**ZUSAMMENFASSUNG:** Es handelt sich nicht nur um eines der faszinierendsten philologischen Probleme der Lyrik des Siglo de Oro, sondern die Frage nach der Autorschaft und Authentizität des poetischen Werks von Fernando de Herrera trennt die Fachwelt in zwei Gruppen, die bis heute unversöhnlich erscheinen. Durch die Anwendung neuer Methoden aus dem Bereich der Digital Humanities, hier der Stilometrie und genauer der Rolling Stylometry (jüngst von Maciej Eder entwickelt) soll die Frage unter neuem Licht betrachtet werden. Dazu wird die These untersucht, ob verschiedene poetische Werke von Herrera möglicherweise von unterschiedlichen, mehreren Personen verfasst wurden. Der in der bisherigen Forschung als wahrscheinlichste Kandidat für ein solches „Ghostwriting“ ist der Maler und Herausgeber Francisco de Pacheco.

**PALABRAS CLAVE:** estilometría; Herrera, Fernando de; poesía del Siglo de Oro; humanidades digitales; estilística computacional

**SCHLAGWÖRTER:** Stilometrie; Herrera, Fernando de; Siglo de Oro; Digital Humanities

### 1. El drama textual herreriano

El problema textual y de autoría que afecta a la obra poética de Fernando de Herrera (1534–1597) es uno de los más apasionantes de la poesía del Siglo de Oro. A las dificultades de transmisión textual de su poesía, particularmente

---

<sup>o</sup> Este trabajo forma parte de un proyecto de tesis doctoral financiado por una ayuda de Formación de Profesorado Universitario (FPU) del Ministerio de Educación y Ciencia español (MECD) y dirigido por el Dr. Juan Montero, a quien agradezco sus orientaciones y consejos.

oscuras dentro de la ya complicada circulación y transmisión que caracteriza a los textos poéticos áureos, tenemos que sumar el intenso y prolongado debate crítico sobre la autenticidad de parte de su obra y la importancia de la poesía herreriana dentro del panorama lírico español del XVI, en el que, si seguimos a Dámaso Alonso y Antonio Vilanova, esta sería fundamental para entender la evolución del lenguaje poético desde Garcilaso a Góngora.<sup>1</sup> Los textos del poeta áureo más sobresaliente de la llamada escuela poética sevillana y conocido con el sobrenombre de “El Divino” se nos han transmitido a través de dos ediciones impresas y varios manuscritos. Realizamos a continuación un somero estado de la cuestión de los textos poéticos herrerianos y de la polémica conocida como el “drama textual” herreriano.

La primera edición, *Algunas obras*, conocida como *H*, fue publicada en 1582 y ha sido considerada tradicionalmente como la más autorizada. Con 91 poemas, constituye una selección de la poesía herreriana preparada por el propio Herrera. Tenemos constancia, además, de que el propio poeta participó en el proceso de impresión, revisando pruebas de imprenta. Por todo ello, *H* es considerada prácticamente como un autógrafo.

La polémica y el debate crítico se centran en la segunda edición, *Versos de Fernando de Herrera*, publicada de forma póstuma en 1619 y preparada por el pintor Francisco Pacheco (1564–1644), conocida como *P*. Esta contiene 365 poemas supuestamente herrerianos, de los cuales 88 ya aparecían en *H* y 70 en otros manuscritos de poesía herreriana. De estos la inmensa mayoría han sufrido correcciones y modificaciones que en algunos casos se limitan a palabras concretas y en otros se trata de una auténtica reescritura de los poemas. Esta edición incluye también 205 poemas nuevos de Herrera, no conocidos hasta la fecha ni transmitidos a través de otras fuentes.

Además de estas dos ediciones, nos ha llegado un corpus de poemas sueltos herrerianos, de los cuales destacan los incluidos en el manuscrito conservado en la Biblioteca Nacional de España, *Cisnes del Betis*, publicado por José Manuel Blecua y conocido como *B*. Este contiene 130 poemas de Herrera, 84 de los cuales son comunes a *H* y/o *P*.

Dentro de esta compleja transmisión de las obras poéticas herrerianas, la cuestión textual se ha centrado fundamentalmente en la discusión sobre la autenticidad de *Versos* por su carácter póstumo y por las diferencias que pre-

---

<sup>1</sup> Antonio Vilanova, “Fernando de Herrera”, in *Historia general de las literaturas hispánicas*, II, publ. bajo la dir. de Guillermo Díaz-Plaja (Barcelona: Barna, 1968), 689–751.

senta con *Algunas obras*, así como las oscuras circunstancias que rodean a la edición.

La única información que tenemos sobre su proceso de preparación se encuentra en los textos preliminares, caracterizados por la ambigüedad. En primer lugar, en el prólogo de Francisco Medina se nos habla del olvido que padecían las obras de Herrera, elogiándose la labor de Pacheco en recoger sus obras perdidas, pero no se nos dan más detalles sobre las causas de esta pérdida y olvido. El licenciado Enrique Duarte da algunos detalles más en su prólogo. Declara que Herrera tenía preparado un volumen de sus obras poéticas para darlo a la imprenta cuando falleció y que estas se han podido recuperar del “naufragio” gracias a la diligencia de Pacheco:

I es cierto que su memoria uviera quedado sepultada en perpetuo olvido, si Francisco Pacheco Celebre Pintor de nuestra Ciudad, i afectuoso imitador de sus escritos, no uviera recogido con particular diligencia i cuidado algunos cuadernos i borradores que escaparon d’el naufragio, en que pocos días des- / pues de su muerte perecieron todas sus obras Poeticas; que el tenia corregidas de ultima mano, i encuadradas para darlas a la Empreñta. Dexo en silencio la culpa d’esta perdida, porque soi enemigo de sacar en publico agenas culpas, i juzgo por merecedor de gran premio, al que con tantas veras à procurado restaurarla, hurtando muchas oras de su mas forçosa y precisa ocupacion; porque no solo copiò una i dos vezes de su mano lo que aora nos ofrece, pero cumplió lo que faltaba de otros papeles sueltos, que avian venido a manos de diferentes personas, de quien los uvo; i aunque todo ello sea d’el mesmo Autor es cosa cierta, que lo que el tenia escogido, i perficionado para sacar a luz seria de mayor, i de mas acabada perfeccion.<sup>2</sup>

Aunque estos testimonios han hecho desconfiar ya a algunos críticos, sin duda el que ha causado más interrogantes ha sido el poema del propio Pacheco que abre el libro, en el que dice: “Goza, ô Nacion osada, el don fecundo | que t’ofresco, en la forma verdadera | qu’imaginè, d’el culto i gran Herrera; | i el fruto de su ingenio, alto i profundo.”<sup>3</sup> ¿Cómo tenemos que interpretar este “en la forma verdadera | qu’imaginè”?

Quevedo fue el primero en desconfiar y atacar la edición preparada por Pacheco en el prólogo a sus lectores de su edición de las poesías de Francisco de la Torre (Madrid, 1631). Sin embargo, su objetivo es por un lado demostrar que Herrera imitó en su poesía la de Francisco de la Torre y, por otro,

<sup>2</sup> Fernando de Herrera, *Obra poética*, ed. José Manuel Blecua (Madrid: Boletín de la Real Academia Española, 1975), t. II: 25–6.

<sup>3</sup> Herrera, *Obra poética*, t. II, 31.

de acuerdo con Peter M. Komanecky, criticar la edición de Pacheco por su mayor cercanía al estilo poético de Góngora:

the real point Quevedo was trying to make was not that the poems as they appear in Pacheco's edition were apocryphal [...] but rather that Herrera should not have developed a more 'baroque' style in his later years, and if he did, then he was worthy of the strongest possible censure.<sup>4</sup>

En cualquier caso, la desconfianza hacia *P* aumentó cuando en el siglo xx Adolphe Coster realizó la primera edición crítica de la poesía herreriana en 1908, en cuyo prólogo defiende que los poemas publicados en *P* no tienen carácter auténtico y que *P* recogería poemas deteriorados por copistas y correctores, o rechazados por el propio Herrera, si bien unos años más tarde, en 1914, decide publicar *P* y reconoce el interés de los poemas publicados por Pacheco.<sup>5</sup>

Con la publicación del manuscrito *B* en 1948 por José Manuel Blecua se descubre que muchos poemas de *P* que no aparecían en *H* sí estaban en este manuscrito y se confirma la tendencia de Herrera a corregir sus escritos a la que hacen referencia los amigos:

I esta fue la causa, de que Fernando de Herrera pareciesse tan difícil, i tardo en aprovar las obras, que via, no porque no admirasse las suyas, que de ninguna cosa estaba mas lexos; porque, como a ombre a quien el uso i exercicio de aquellas cosas avia dado una mui entera noticia de los preceitos mas ocultos de l'arte, le satisfazian pocas, i sus oidos como capaces de otras mayores desseavan siempre alguna de consumada perfeccion; de que pueden dar testimonio los borradores de sus Versos, que despues de limados muchas vezes, i en espacio de años enteros, apenas le contentavan; i assi desechó muchos, que pudieran ser estimados de los mas entendidos en esta profession.<sup>6</sup>

A partir de entonces, las diferencias que presenta *P* con respecto a *H* han dividido a la crítica en dos grandes grupos:

Por un lado, el primero se iniciaría con la postura de Adolphe Coster, cuyo principal defensor ha sido el reconocido estudioso José Manuel Blecua, y a la que se han adherido otros críticos como Cristóbal Cuevas, Vicente García de Diego o A. David Kossoff, que niegan la autoría herreriana de *P* y sospechan

<sup>4</sup> Peter Komanecky, "Quevedo's Notes on Herrera: The Involvement of Francisco de La Torre in the Controversy Over Góngora", *Bulletin of Hispanic Studies* 52, no. 2 (1975): 123–33, cit. 125.

<sup>5</sup> Fernando de Herrera, *Algunas obras de Fernando de Herrera*, ed. Adolphe Coster (Paris: Champion, 1908); Fernando de Herrera, *Versos de Fernando de Herrera*, ed. Adolphe Coster (Estrasburgo: Bibliotheca Romanica, 1914).

<sup>6</sup> Herrera, *Obra poética*, t. II, 23.

de una segunda mano, presumiblemente la de Pacheco, el editor, que además de pintor era poeta y gran admirador de Herrera:<sup>7</sup>

[...] solo admitiendo la posibilidad de que Pacheco emendase en la “forma verdadera” que imaginó la obra poética de Herrera tienen respuesta las muchas interrogaciones planteadas a lo largo de estas páginas [...] Su único defecto consistió en que se le fué la mano y restauró a veces con demasiados tintes viejos y nuevos las versiones que le ofrecían los cuadernos, borradores y papeles sueltos remitidos con tanta diligencia.<sup>8</sup>

Por otro, los hispanistas italianos Salvatore Battaglia, Oreste Macrí e Inoria Pepe Sarno defienden la autoría herreriana de *P*. Para ellos, las variantes serían de Herrera y mostrarían una evolución de signo prebarroco.<sup>9</sup>

En un primer momento, la hipótesis defendida por los que negaban la autoría de *P* fue que Pacheco había utilizado borradores de Herrera que recogerían versiones anteriores a los poemas publicados en 1582 y, por tanto, descartadas por El Divino. Sin embargo, Salvatore Battaglia a través del análisis de las variantes de los poemas comunes a *B*, *H* y *P* consiguió convencer a casi la totalidad de la crítica de que el orden de los poemas es  $B \rightarrow H \rightarrow P$ ,<sup>10</sup> y a día de hoy solo el hispanista Ricardo Senabre sigue defendiendo que *P* recoge borradores primitivos.<sup>11</sup> Consecuentemente, el debate se ha centrado desde esta fecha definitivamente en la autenticidad de *P*. En los años 50, José Manuel Blecua y Oreste Macrí publicaron diversos artículos intentando cada uno convencer al otro de su postura sin éxito. Los editores modernos de la

<sup>7</sup> David Kossoff, “Another Herrera Autograph: Two Variant Sonnets”, *Hispanic Review* 33, no. 3 (1965): 318–25; David Kossoff, “Herrera, editor de un poema”, in *Homenaje a Rodríguez-Moñino* (Madrid: Castalia, 1966), 283–90; David Kossoff, “Algunas variantes de versos de Herrera”, *Nueva revista de filología hispánica* XI (1957): 57–63; David Kossoff, “Algo más sobre largo y luengo en Herrera”, *Revista de filología española* XLI (1957): 401–10; David Kossoff, *Vocabulario de la obra poética de Herrera* (Madrid: RAE, 1966); José Manuel Blecua, “De nuevo sobre los textos poéticos de Herrera”, *Boletín de La Real Academia Española* XXXVIII (1958): 377–408; José Manuel Blecua, “Los textos poéticos de Fernando de Herrera”, *Archivum* IV (1954): 247–63; Cristóbal Cuevas, “La cuestión textual”, in *Poesía castellana original completa* (Madrid: Cátedra, 1985), 87–99.

<sup>8</sup> José Manuel Blecua, “De nuevo sobre los textos poéticos de Herrera”, *Boletín de La Real Academia Española* XXXVIII (1958): 406–7.

<sup>9</sup> El profesor Oreste Macrí ha sido quien más ha estudiado esta supuesta evolución de la poesía herreriana con diversos artículos que fueron en aumento y se convirtieron finalmente en la monografía Oreste Macrí, *Fernando de Herrera* (Madrid: Gredos, 1972).

<sup>10</sup> Salvatore Battaglia, “Per il testo di Fernando de Herrera”, *Filologia Romanza*, no. I (1954): 51–88.

<sup>11</sup> Ricardo Senabre, “Los textos emendados de Herrera”, *Edad de Oro* IV (1985): 179–93.

poesía de Herrera han seguido las pautas de Blecua, tomando como modelo su edición de la poesía herreriana.<sup>12</sup>

En los últimos años, algunos estudiosos entre los que destacan Inoria Pepe Sarno y Juan Montero han entrevisto la posibilidad de realizar avances en el asunto aplicando la metodología utilizada por Battaglia, esto es, el análisis crítico de las variantes entre los diferentes testimonios a textos sueltos y, en ocasiones, desconocidos de El Divino, que han supuesto interesantes aportaciones.<sup>13</sup>

Sin embargo, parece que no se puede llegar a conclusiones definitivas con la metodología adoptada hasta ahora, por lo que se hace necesario un nuevo enfoque que abarque todo el corpus y que aproveche las posibilidades que ofrecen hoy las Humanidades Digitales y, más concretamente, la Estilometría.<sup>14</sup>

## 2. La Estilometría<sup>15</sup>

La Estilometría, también llamada Estilística Computacional, es una de las ramas más destacadas a día de hoy dentro de las Humanidades Digitales. Sus

<sup>12</sup> Herrera, *Obra poética*.

<sup>13</sup> Juan Montero, “Una versión inédita (con algunas variantes) de la canción ‘Al sueño’ de F. Herrera”, *Cuadernos de investigación filológica XII–XIII* (1987): 117–32; Juan Montero, “Dos textos poéticos de Fernando de Herrera con variantes y un posible soneto desconocido (más una lira antiherreriana)”, *Bulletin hispanique* 107, no. 2 (2005): 605–16; Inoria Pepe Sarno, “La ‘luz’ de la ‘Aurora’: variantes en dos sonetos de Fernando de Herrera”, in *AIH. VIII Congreso*, 1983; Inoria Pepe Sarno, “Marte-Don Juan in una canzone di Fernando de Herrera”, *Archivo hispalense*, 1986; Inoria Pepe Sarno, “Bianco il ghiaccio, non il velo: ritocchi e metamorfosi in un sonetto di herrera”, *Strumenti Critici*, no. XLVI (1981): 458–71; Inoria Pepe Sarno, “Il senso della storia in alcuni sonetti di Fernando de Herrera”, in *Estudios en homenaje a D. Claudio Sánchez Albornoz en sus 90 años* (Ávila-Buenos Aires: Fundación “Claudio Sánchez Albornoz/Instituto de Historia de España”, 1990), 307–19.

<sup>14</sup> Existen ya un par de estudios que aplican enfoques de corpus o computacionales a la obra de Herrera, aunque todavía no a la cuestión del drama textual: E. Espinilla Buisán et al., *Garcilaso y Herrera: vocabulario, frecuencias y concordancias de las anotaciones a la poesía de Garcilaso de Fernando de Herrera* (Barcelona: Edicions de la Universitat de Barcelona, 2003); Laura Hernández Lorenzo, “The Poetic Word of Fernando de Herrera: An Approach through Corpus and Computational Linguistics”, in *CILC 2016. 8<sup>th</sup> International Conference on Corpus Linguistics: EPiC Series in Language and Linguistics*, ed. Antonio Moreno Ortiz and Chantal Pérez-Hernández, vol. 1, 2016, 170–80.

<sup>15</sup> Un estado de la cuestión en español bastante completo en cuanto a los orígenes de la Estilometría lo encontramos en José Manuel Fradejas Rueda, “El análisis estilométrico aplicado a la literatura española: las novelas policíacas históricas”, *Caracteres: estudios culturales y críticos de la esfera digital* 5, no. 2 (2016): 196–246. En este hemos basado el nuestro y a él remitimos para más información.

comienzos se remontan a la llamada *Shakespeare Authorship Question* o la cuestión de si William Shakespeare escribió todo lo que se le atribuye. Los investigadores Augustus de Morgan (1806–1871) y Thomas C. Mendenhall (1841–1924) compararon la longitud de las palabras de las obras atribuidas a Shakespeare con las de otros candidatos como Vere, Bacon y Marlowe. Se trataba, sin embargo, de un procedimiento muy mecánico e insuficiente para llegar a conclusiones sólidas sobre la autoría de un texto. El primer gran éxito fue el caso de los *Federalist Papers* en el que los investigadores Mostaller y Wallace utilizaron por primera vez la frecuencia de las palabras gramaticales o palabras función (*function words*) y consiguieron establecer la autoría de doce de los artículos, reclamada tanto por Hamilton como por Madison. Desde entonces, se han realizado numerosos estudios de autoría utilizando estudios estilométricos, siendo uno de los más sonados el caso de *The Cuckoo's Calling*, firmada por un tal Robert Galbraith, que resultó ser un pseudónimo de J. K. Rowling, tal y como Patrick Juola había obtenido en todos sus análisis estilométricos.<sup>16</sup>

En el caso de la literatura española, sin embargo, los estudios estilométricos son aún bastante escasos, sobre todo si los comparamos con los aplicados a textos en inglés.<sup>17</sup> Entre ellos destaca la aplicación de la Estilometría a las dos obras anónimas más famosas de la literatura española, el *Lazarillo de Tormes* y el *Quijote* de Avellaneda. Aunque no se ha podido determinar la autoría de estas obras, en el caso del *Lazarillo* sí se han podido descartar candidatos<sup>18</sup> y en el del *Quijote* de Avellaneda ha quedado resaltada una, por otra parte, no del todo sorprendente cercanía de esta obra con las del propio Cervantes.<sup>19</sup>

<sup>16</sup> Patrick Juola, “The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions”, *Digital Scholarship in the Humanities* 30, no. 1 (2015): 100–13.

<sup>17</sup> Esta situación ha llevado a la aparición de artículos divulgativos en español sobre los métodos estilométricos como el ya mencionado Fradejas Rueda, “El análisis estilométrico aplicado a la literatura española: las novelas policíacas históricas”; y otros como José Calvo Tello, “Entendiendo delta desde las humanidades”, *Caracteres: estudios culturales y críticos de la esfera digital* 5, no. 1 (2016): 140–76.

<sup>18</sup> Nanette Rißler-Pipka, “Digital Humanities und die romanische Literaturwissenschaft: der Autorschaftsstreit um den Lazarillo de Tormes”, *Romanische Forschungen* 128, no. 3 (2016): 316–42; Javier de la Rosa and Juan Luis Suárez, “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 (2016): 373–438.

<sup>19</sup> Javier Blasco, “Avellaneda desde la estilometría”, in *Cervantes: los viajes y los días*, ed. Pedro Ruiz Pérez (Madrid: Prosa Barroca y SIAL Ediciones, 2016), 97–116; Nanette Rißler-Pipka, “Avellaneda y los problemas de la identificación del autor: propuestas para una investigación con nuevas herramientas digitales”, in *El otro Don Quijote: la continuación de Fernández de Avellaneda*

Por otra parte, en varios congresos y reuniones científicas he podido conocer otros trabajos estilométricos aplicados a la literatura española que se están llevando a cabo sobre textos medievales castellanos, el teatro barroco o la novela de la Edad de Plata. Por lo general, la poesía lírica suele ser el género literario menos abordado a través de métodos computacionales debido a la mayor presencia de subjetividad y este es también el caso en la aplicación de la Estilometría a la literatura española, aunque destaca el análisis que está acometiendo la profesora Nanette Rißler-Pipka con la poesía de Góngora y Picasso y el publicado por Antonio Rojas Castro sobre la fábula mitológica del XVII.<sup>20</sup>

### 3. Preparación de los textos

Para aplicar los métodos computacionales y estilométricos al problema textual de autoría de la poesía herreriana resultaba necesario en primer lugar contar con los textos de Herrera en formato digital. Puesto que la obra poética solo se había digitalizado de forma muy parcial en la web, el primer paso necesario fue acometer la digitalización de esta en su totalidad, la cual se llevó a cabo aplicando el software de OCR (Optical Character Recognition) Adobe Acrobat Pro sobre la edición más autorizada, la de José Manuel Blecua, que contiene además la totalidad de los textos poéticos herrerianos.<sup>21</sup> A continuación, se revisó el corpus y se corrigieron los errores del OCR. Debido a que las particularidades ortográficas de Herrera hubieran entorpecido el análisis estilométrico, aunque estas se mantuvieron por su interés como otro indicador de cambio entre los textos herrerianos, se creó un corpus paralelo en el que esta fue modernizada de acuerdo con las reglas actuales de la Real Academia de la Lengua.

Por otra parte, Francisco Pacheco, el editor de *P*, es el principal sospechoso de los críticos que niegan la autoría herreriana de esta obra, que piensan que pudo haber retocado los poemas para adaptarlos al gusto más barroco del momento de publicación:

Sin embargo, los textos de Pacheco presentan la mixtura más informe y extraña que conoce la poesía española del siglo XVI, y sólo un editor que en 1619 (en pleno furor culterano) quiere dar un tinte nuevo y viejo a la vez a la poesía,

---

y sus efectos, ed. Hanno Ehrlicher (Augsburgo, 2016), 27–51.

<sup>20</sup> Antonio Rojas Castro, “Luis de Góngora y la fábula mitológica del Siglo de Oro: clasificación de textos y análisis léxico con métodos informáticos”, *Studia Aurea* 11 (2017): 111–42.

<sup>21</sup> Herrera, *Obra poética*.

“en la forma verdadera que imaginó”, puede atreverse a cambiar sistemáticamente el pronombre os en vos [...].<sup>22</sup>

Tomando esto en consideración, hemos recopilado su poesía de diferentes fuentes, principalmente del *Arte de la pintura* y del *Libro de los retratos*, junto con la *Biblioteca de autores españoles*, y los poemas recogidos por Asensio.<sup>23</sup> Este corpus fue también modernizado siguiendo los mismos criterios que con la poesía herreriana.<sup>24</sup>

#### 4. Análisis estilométrico

Como hemos visto, el problema de autoría de *P* tiene una gran complejidad, ya que no se trata solo de confirmar si la obra es de Herrera o no lo es, sino más bien de detectar si hay otra mano aparte de la de El Divino, que indudablemente está en las partes coincidentes con *B* y *H* que no han sido modificadas.

En consecuencia, este primer acercamiento estilométrico a los problemas de autoría en la poesía de Herrera se basa en uno de los métodos desarrollados para analizar trabajos colaborativos u obras en las que intervienen dos o más manos, Rolling Stylometry, presentado por Maciej Eder en un artículo publicado en 2016 en *Digital Scholarship in the Humanities (DSIH)*, la revista de la Alliance of Digital Humanities Organisations (ADHO).<sup>25</sup> En este trabajo, Eder

<sup>22</sup> Blecua, “De nuevo sobre los textos poéticos de Herrera”, 396.

<sup>23</sup> José María Asensio, *Francisco Pacheco: sus obras artísticas y literarias. introducción e historia del libro de descripción de verdaderos retratos de ilustres y memorables varones que dejó inédito* (Sevilla: Rasco, 1886), esta obra se ha consultado a través de la Biblioteca Virtual de Andalucía en el siguiente enlace: [www.bibliotecavirtualdeandalucia.es/catalogo/consulta/registro.cmd?id=1017777](http://www.bibliotecavirtualdeandalucia.es/catalogo/consulta/registro.cmd?id=1017777), consulta: 14/02/2018; Adolfo de Castro, ed., *Biblioteca de autores españoles, desde la formación del lenguaje hasta nuestros días: poetas líricos de los siglos XVI y XVII* (Madrid: Rivadeneyra, 1854), esta obra se ha consultado a través del repositorio archive.org en el siguiente enlace: <https://archive.org/details/bibliotecadeauto32madruoft>, consulta: 14/02/2018; Francisco Pacheco, *Libro de descripción de verdaderos retratos de ilustres y memorables varones* (Sevilla: Rafael Tarasco, 1881), esta obra se ha consultado a través de la Biblioteca Virtual de Andalucía en el siguiente enlace: [www.bibliotecavirtualdeandalucia.es/catalogo/consulta/registro.cmd?id=1040827](http://www.bibliotecavirtualdeandalucia.es/catalogo/consulta/registro.cmd?id=1040827), consulta: 14/02/2018; Francisco Pacheco, *Arte de la pintura ...* (Madrid: Galiano, 1866), esta obra se ha consultado a través de Google Books en el siguiente enlace: [https://books.google.de/books/about/Arte\\_de\\_la\\_pintura.html?id=litPAAAYAAJ&redir\\_esc=y](https://books.google.de/books/about/Arte_de_la_pintura.html?id=litPAAAYAAJ&redir_esc=y), consulta: 14/02/2018.

<sup>24</sup> A la hora de preparar esta recopilación, me han resultado de enorme utilidad los materiales que muy generosamente me ha cedido el Dr. Juan Montero, a quien se lo agradezco mucho.

<sup>25</sup> Maciej Eder, “Rolling Stylometry”, *Digital Scholarship in the Humanities* 31, no. 3 (2016): 457–69.

aplica tres técnicas de Machine Learning (Machine Learning classifiers) cuya eficacia ha sido probada a través del paquete *stylo* de R:<sup>26</sup> support vector machines (svm), nearest shrunken centroids (NSC) y la Delta de Burrows.<sup>27</sup> Con estas aborda tres casos de obras con varias manos: el *Roman de la Rose*, poema del medievo francés escrito por Guillaume de Lorris y Jean de Meun; la Biblia de la Reina Sofía, obra con un complicado problema de transmisión en la que han participado varios traductores; y los trabajos colaborativos de Joseph Conrad y Ford Madox Ford. En todos los casos, las señales autoriales se detectan con buenos resultados.

Una versión anterior de Rolling Stylometry, Rolling Delta, que aplica solo la medida de Burrows, había sido aplicada anteriormente a los trabajos colaborativos de Conrad y Ford por los investigadores Jan Rybicki, David Hoover y Mike Kestemont<sup>28</sup> y a los trabajos colaborativos de Charles Dickens por el profesor Tomoji Tabata.<sup>29</sup> Por otra parte, tras la publicación del artículo de Rolling Stylometry, han aparecido más trabajos que aplican esta técnica a otras obras en inglés.<sup>30</sup>

Previamente al análisis con los textos herrerianos, se decidió realizar primero una prueba con un texto de la literatura en español con dos manos y en el que supiéramos antes de realizar el análisis qué parte fue escrita por cada autor. El objetivo era confirmar la eficacia de este procedimiento con textos literarios en español. A tal efecto se aplicó la técnica sobre la obra barroca *Amor es más laberinto*, co-escrita por Sor Juana Inés de la Cruz y Fray Juan de Guevara.<sup>31</sup> Las marcas ‘a’ y ‘b’ marcan las separaciones en jornadas de la obra.

<sup>26</sup> Maciej Eder, Jan Rybicki, and Mike Kestemont, “Stylometry with R: A Package for Computational Text Analysis”, *The R Journal* XX (2008): 1–15.

<sup>27</sup> John Burrows, “Delta: a Measure of Stylistic Difference and a Guide to Likely Authorship”, *Literary and Linguistic Computing* 17, no. 3 (2002): 267–87.

<sup>28</sup> Jan Rybicki, David Hoover, and Mike Kestemont, “Collaborative Authorship: Conrad, Ford and Rolling Delta”, *Literary and Linguistic Computing* 29, no. 3 (2014): 422–31.

<sup>29</sup> Tomoji Tabata, “Stylometry of Collaborations: Dickens, Collins and Their Collaborative Writings”; in *Digital Humanities 2014: Conference Abstracts* (Lausanne: EPFL-UNIL, 2014), 378–80.

<sup>30</sup> Hartmut Ilsemann, “More News on Sir Thomas More”, *Digital Scholarship in the Humanities*, 2017, 1–13; Alexander A.G. Gladwin, Matthew J. Lavin, and Daniel M. Look, “Stylometry and Collaborative Authorship: Eddy, Lovecraft, and ‘The Loved Dead’”, *Digital Scholarship in the Humanities* 32, no. 1 (2017): 123–40.

<sup>31</sup> La colaboración en la escritura de esta obra dramática consta desde la edición princeps de 1692, leyéndose en la portada de la comedia lo siguiente: “AMOR ES MAS LABERYNTO, | COMEDIA, | DE LA QVAL LAS JORNADAS PRIMERA, | y tercera son de la Madre Juana, y la segunda del /Lic. D. Juan de Guevara, Ingenio Conocido | de la Ciudad de Mexico” [folio 389]. La edición puede consultarse a través de la Biblioteca Virtual Miguel de Cervantes en el siguiente enla-

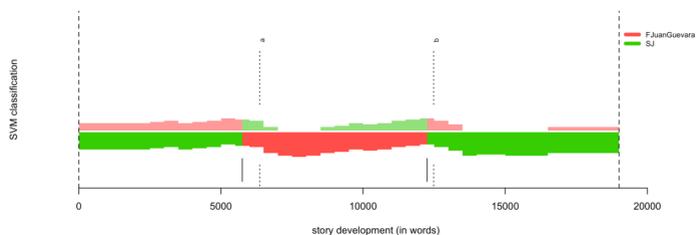


Fig. 1: *Amor es más laberinto*. SVM

Como puede observarse en la figura 1, los resultados reflejan lo que ya sabíamos, que el primer y el tercer acto fueron escritos por Sor Juana y el segundo por Fray Juan de Guevara. La probabilidad de autoría es incluso del 100 % en algunas partes del gráfico.

En este análisis estilométrico de la poesía herreriana, se aplicarán las tres técnicas incluidas en Rolling Stylometry a través del paquete *stylo* en R y utilizando como corpus de referencia por un lado, las obras poéticas que sabemos indudablemente que pertenecen a Herrera, es decir, los poemas sueltos y *H*; y, por otro, los textos poéticos que hemos recopilado de Pacheco. El ‘test corpus’ será tanto *P* como un corpus que hemos preparado con los 205 poemas nuevos de esta edición, al que hemos llamado *P2*. Nuestro objetivo es medir de forma más científica y objetiva la probabilidad de la autoría herreriana de *P*, así como rechazar o confirmar en la medida de lo posible la existencia de una segunda mano y más concretamente la de Francisco Pacheco.

Comenzamos utilizando Rolling SVM y en el primer experimento utilizaremos las mismas variables que utiliza Maciej Eder en su artículo, esto es, muestras de 5000 palabras con solapamiento de 4500. Además, en la visualización de los resultados la separación entre libros de *P* se muestra a través de las marcas ‘a’ y ‘b’.

Con Rolling SVM y las 100 palabras más frecuentes sobre *P* y *P2*, obtenemos las figuras 2 y 3, en las que, como puede verse, la probabilidad de mano de Herrera es casi del 100 %.

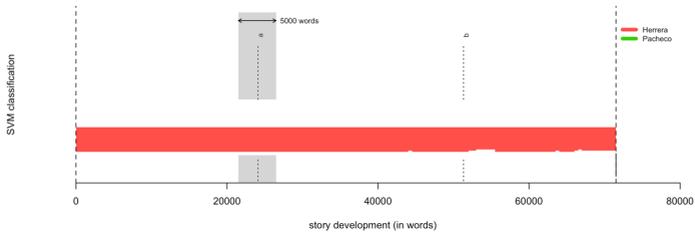


Fig. 2: P. SVM. Samples of 5000 words. 100 MFW

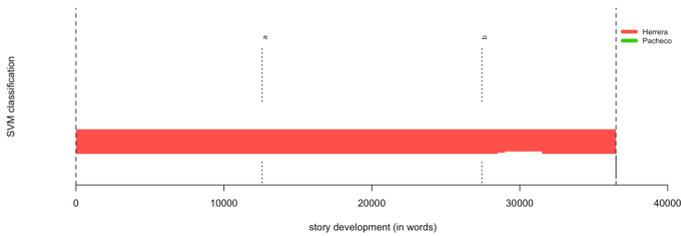


Fig. 3: P2. SVM. Samples of 5000 words. 100 MFW

En la segunda prueba que hemos realizado con SVM hemos utilizado muestras de 3000 palabras con un solapamiento de 2500. Obtenemos las figuras 4 y 5, en las que aparece algo de probabilidad (si bien ciertamente absolutamente mínima) de presencia de mano de Pacheco al comienzo del tercer libro de P.

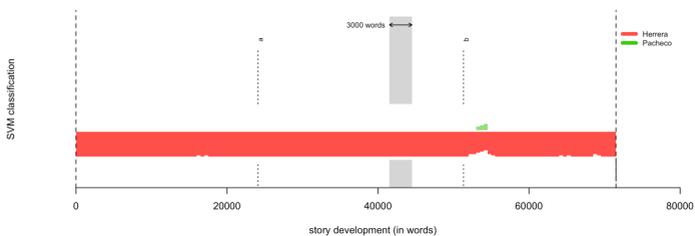


Fig. 4: P. SVM. Samples of 3000 words. 100 MFW

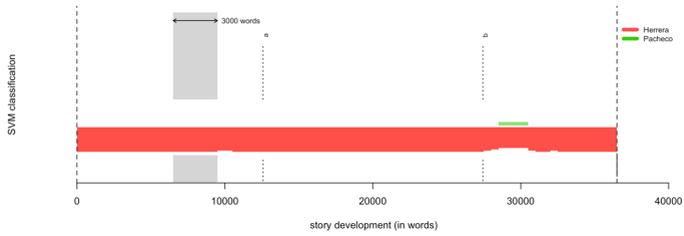


Fig. 5: P2. SVM. Samples of 3000 words. 100 MFW

A continuación, aplicamos NSC con las 50 palabras más frecuentes sobre *P* y *P2* la elección es Herrera para todo el texto en ambos casos, aunque aparece una probabilidad mínima de Pacheco en algunas partes del texto, como puede verse en las figuras 6 y 7.

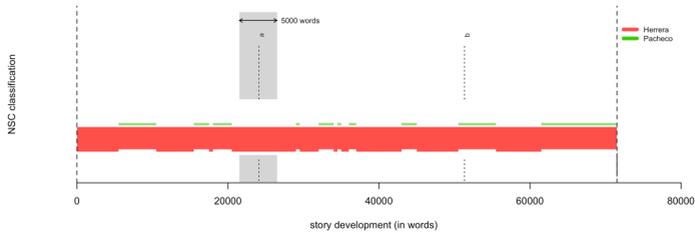


Fig. 6: P. NSC. Samples of 5000 words

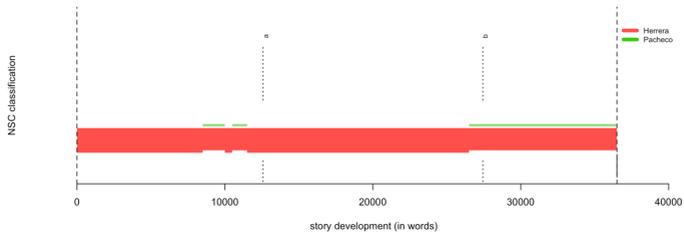


Fig. 7: P2. NSC. Samples of 5000 words

En tercer lugar, aplicamos Rolling Delta con las 500 palabras más frecuentes sobre  $P$  y  $P_2$ . Como puede verse en las figuras 8 y 9, la primera elección es Herrera y la segunda Pacheco.

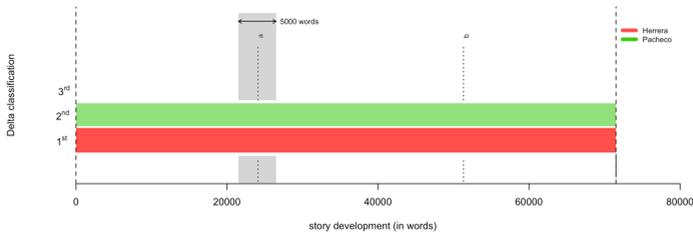


Fig. 8: P.Delta. Samples of 5000 words

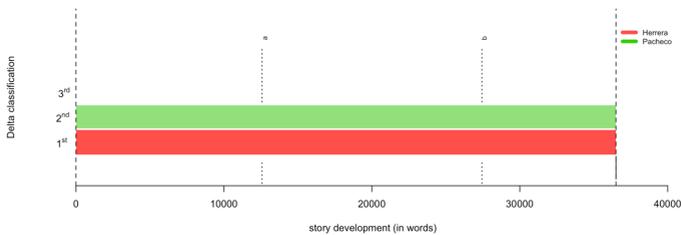


Fig. 9: P2. Samples of 5000 words

Hasta ahora mediante todas las técnicas empleadas el resultado para todo  $P$  y  $P_2$  es Herrera y la probabilidad de presencia de mano de Pacheco parece ser casi inexistente. Para confirmar este resultado y asegurarnos de que las diferencias de tamaño no estuviesen influyendo en el mismo –puesto que el corpus de Herrera es mucho mayor que el de Pacheco, 77044 frente a 5085 palabras–, decidimos realizar un nuevo análisis reduciendo previamente el corpus de referencia de Herrera y  $P$  a 8000 palabras aproximadamente cada uno.

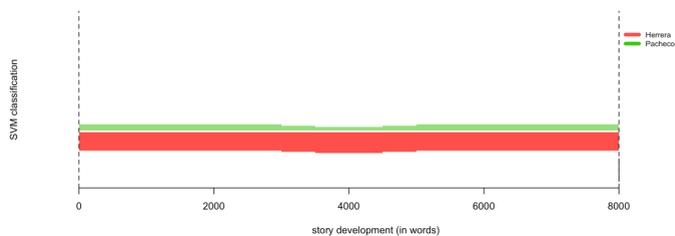


Fig. 10: SVM. Samples of 5000 words. 100 MFW. Aleatory reduced version P

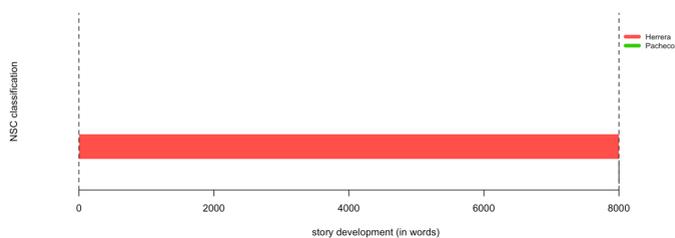


Fig. 11: NSC. Samples of 5000 words. Aleatory reduced version P

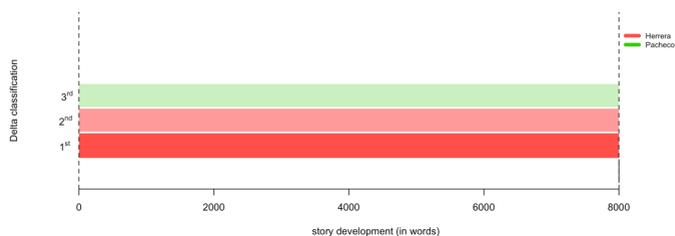


Fig. 12: Delta. Samples of 5000 words. Aleatory reduced version P

Como puede verse en las figuras 10, 11 y 12, aunque la probabilidad de presencia de Pacheco ha aumentado en SVM, en todas las opciones la autoría elegida a lo largo de todo el texto es la de Herrera, no eligiéndose la de Pacheco en ningún caso o fragmento.

## 5. Conclusiones

Con este trabajo se pretende mostrar en un primer nivel la utilidad de las herramientas digitales y, especialmente, de las estilométricas para avanzar en la investigación del drama textual herreriano; y, en un segundo nivel, el interés por supuesto de aplicar la Estilometría a textos en español, como han hecho algunos trabajos pioneros, y de utilizar técnicas como Rolling Stylometry para abordar obras literarias en español con autoría mixta o que pueden tener varias manos, un campo que está todavía casi sin explorar.

Centrándonos en los resultados de nuestro estudio, aparentemente en este primer acercamiento estilométrico parece que la probabilidad de que hubiera mano de Pacheco en *P* sería muy baja, aunque habría que tener en cuenta por un lado, que en el corpus de poesía de Pacheco encontramos temas no presentes en *P* (como la pintura), y, por otro lado, que los resultados podrían variar si obtuviéramos más poemas del autor del *Libro de retratos* para añadir al corpus de referencia. En cualquier caso, con los datos que tenemos a día de hoy el argumento que esgrimen los detractores de la autoría herreriana de que con bastante seguridad la mano del pintor y editor literario sería la responsable de los cambios en *P* en su intento de acercar el estilo herreriano al gusto más barroco de la época pierde fuerza.

Para concluir, nos gustaría señalar que este estudio forma parte de un proyecto mayor en desarrollo y que será necesaria la realización de nuevos análisis y la aplicación de nuevas técnicas estilométricas con un corpus mayor de autores para poder llegar a conclusiones definitivas sobre la autoría herreriana de *P*.