

“Reis melhor do que eu”: los heterónimos de Pessoa desde una perspectiva estilométrica

Daniil Skorinkin

Universidad de Potsdam, Potsdam, Alemania

skorinkin.danil@gmail.com

Boris Orekhov

Universidad Nacional de Investigación “Escuela Superior de Economía”, Moscú, Rusia

Instituto de Literatura Rusa de la Academia Rusa de Ciencias, San Petersburg, Rusia

nevmenandr@gmail.com

Tradicionalmente, la estilometría se ha utilizado para resolver problemas de atribución de autoría. Los métodos cuantitativos de atribución siguen siendo la última herramienta de los investigadores cuando no se dispone de pruebas documentales fiables. En los últimos veinte años, el método Delta, desarrollado por John F. Burrows, se ha impuesto como el principal método de atribución. En general, ha demostrado ser una forma bastante fiable de atribuir textos en casos controvertidos. Sin embargo, como muestra nuestra investigación, destaca el caso de Fernando Pessoa, quien produjo sus textos “en nombre” de identidades ficticias, comúnmente llamadas “heterónimos”. Delta no identificó dichas obras como se esperaba, es decir, como textos pertenecientes a la pluma de una sola persona, Fernando Pessoa, sino como textos de diferentes autores. El artículo lleva a cabo una serie de experimentos para comprobar hasta qué punto Pessoa consigue confundir la valoración cuantitativa de autoría de sus textos poéticos. Los textos de Pessoa son examinados como un corpus independiente y con el trasfondo de la obra de otros poetas lusófonos. En todos los casos, las distancias entre los textos pertenecientes a los heterónimos de Pessoa son comparables a las distancias entre textos de autores diferentes, es decir, mucho mayores que las distancias entre los textos de un mismo autor.

Palabras clave: atribución de autoría; Delta; estilometría; Fernando Pessoa; heterónimos; métodos cuantitativos en humanidades.

Cómo citar este artículo (MLA): Skorinkin Daniil y Boris Orekhov. “Reis melhor do que eu”: los heterónimos de Pessoa desde una perspectiva estilométrica. *Literatura: teoría, historia, crítica*, vol. 25, núm. 2, 2023, págs. 155-179

Artículo original. Recibido: 15/01/23; aceptado: 12/04/2023. Publicado en línea: 01/07/2023.



“Reis melhor do que eu”: Pessoa’s Heteronyms from a Stylometric Perspective

Traditionally, stylometry has been used to solve problems of authorship attribution. Quantitative attribution methods remain the last hope for researchers when reliable documentary evidence is unavailable. In the last 20 years, the Delta method, developed by John F. Burrows, has emerged as the leading attribution method. Overall, it has proven to be a reasonably reliable way of attributing texts in controversial cases. However, as our research shows, the case of Fernando Pessoa stands out, as he produced his texts “on behalf” of fictitious identities, commonly known as “heteronyms”. It turns out that Delta does not identify these works as expected, that is, as texts belonging to the pen of a single person, Fernando Pessoa, but as texts from different authors. The article carries out a series of experiments to test the extent to which Pessoa manages to confuse the quantitative assessment of the authorship of his poetic texts. Pessoa’s texts are examined as an independent corpus and against the background of the work of other Lusophone poets. In all cases, the distances between texts belonging to Pessoa’s heteronyms are comparable to those between texts from different authors, much greater than the distances between texts from the same author.

Keywords: authorship attribution; Delta; stylometry; Fernando Pessoa; heteronyms; quantitative methods in humanities.

“Reis melhor do que eu”: os heterônimos de Pessoa sob uma perspectiva estilométrica

Tradicionalmente, a estilometria tem sido usada para resolver problemas de atribuição de autoria. Os métodos de atribuição quantitativa continuam sendo a última esperança dos pesquisadores quando não há evidências documentais confiáveis disponíveis. Nos últimos vinte anos, o método Delta, desenvolvido por John F. Burrows, estabeleceu-se como o principal método de atribuição. Em geral, ele provou ser uma maneira bastante confiável de atribuir textos em casos controversos. No entanto, como mostra nossa pesquisa, destaca-se o caso de Fernando Pessoa, que produziu seus textos “em nome” de identidades fictícias, comumente chamadas de “heterônimos”. Como o artigo mostra, o Delta não identifica essas obras como esperado, ou seja, como textos pertencentes à pena de uma única pessoa, Fernando Pessoa, mas como textos de diferentes autores. O artigo realiza uma série de experimentos para testar até que ponto Pessoa consegue confundir a avaliação quantitativa da autoria de seus textos poéticos. Os textos de Pessoa são examinados como um *corpus* independente e contra o pano de fundo do trabalho de outros poetas lusófonos. Em todos os casos, as distâncias entre os textos pertencentes aos heterônimos de Pessoa são comparáveis às distâncias entre textos de diferentes autores, ou seja, muito maiores do que as distâncias entre textos do mesmo autor.

Palavras-chave: estilometria; Fernando Pessoa; heterônimos; atribuição de autoria; Delta; métodos quantitativos em Ciências Humanas.

Introducción

CUANDO SE MENCIONA LA ESTILOMETRÍA, lo primero en lo que suelen pensar los investigadores es en la atribución de autoría con métodos cuantitativos. Efectivamente, la principal aplicación de la estilometría se encuentra precisamente en que es una forma de establecer la autoría de un texto, guiándose no por testimonios documentales, sino por cálculos que se pueden realizar sobre el texto mismo (Mosteller y Wallace; Forsyth y Holmes). Estos trabajos demuestran que la medición de los rasgos estilísticos permite distinguir lo que comúnmente se denomina una señal autoral (*author signal*). Desde que John F. Burrows introdujo su método Delta, su enfoque ha dominado el campo y ha sido validado para diferentes lenguas y géneros (Eder y Rybicki; Eder, Rybicki y Kestemont; García y Calle-Martín; van Dalen-Oskam y van Zundert).

Se trata de un enfoque cuantitativo, basado en la metodología de comparar textos de un corpus de investigación por parejas y calcular la menor distancia entre los textos atribuidos con precisión y los textos de cuya autoría dudamos. En la mayoría de los casos, esta distancia mínima significa que los textos son escritos por el mismo autor.

La estilometría se compara a menudo con la dactiloscopia, porque se utiliza para buscar una especie de “huella dactilar” del autor, un trazo individual y único en un texto. Sin embargo, ya se ha demostrado repetidamente que esto no es más que una metáfora. Ni Delta ni otros métodos no tan extendidos permiten identificar una verdadera “huella dactilar” en un texto. Como está bien explicado en el trabajo de Jannidis y Lauer, Delta simplemente muestra las relaciones de un texto con otros dentro de algún conjunto de obras (29-54). Si este conjunto cambia, también cambiarán las relaciones dentro del conjunto, lo que posiblemente conducirá a una agrupación diferente de los textos y, a veces, incluso a conclusiones distintas sobre la cuestión de la autoría.¹ Adicionalmente, la agrupación puede verse afectada por muchos factores. Ya se ha demostrado que los resultados de la atribución a través del

1 Vale la pena aclarar, sin embargo, que la composición del corpus no necesariamente lleva a conclusiones diferentes sobre la autoría, sino solamente en casos complejos.

método de Delta de Burrow pueden verse afectados por el género literario,² la evolución del estilo del autor de las primeras a las últimas obras,³ el estilo del traductor,⁴ la edición y corrección editorial⁵ o el estilo del habla de los personajes.⁶

En esta lista, llaman especial atención los idiolectos de los personajes. La caracterización del habla de los personajes literarios es una manera de crear una “voz”, que se definiría como especial, distinto del resto del texto o, al menos, de la forma de hablar de otros personajes. El autor aborda conscientemente la creación de esa voz, tratando de dotarla de rasgos individualizados. Se sabe que, cuando el investigador dispone de datos suficientes, los personajes generan una fuerte señal estilométrica que, en efecto, difiere del estilo del autor y del estilo de habla de otros personajes. Así, los grupos de personajes similares en la Trilogía de Henryk Sienkiewicz se distinguen estilométricamente (Rybicki, “Burrowing”). Un trabajo de Daniil Skorinkin muestra que los personajes principales de la epopeya *Guerra y paz* de Tolstoi hablan con voces distintivas, y que fragmentos de discurso aleatorios, pertenecientes al mismo personaje, se terminan agrupados (55-70). Lo mismo se ha demostrado estadísticamente para los diálogos de Platón (Sansone). Estos ejemplos sugieren que, en principio, el autor es capaz de crear a voluntad idiolectos estilísticamente diferentes y que este fenómeno puede ser capturado con métodos estilométricos. Al mismo tiempo, como fue demostrado para el drama de la Edad Moderna temprana, la diferencia entre los personajes no es tan grande como para destruir la unidad del estilo autoral (Burrows y Craig). En otras palabras, aunque los personajes difieran entre sí en cuanto al estilo, la diferencia estilométrica entre distintos autores es mayor.

Hasta ahora no se ha cuestionado que la creación de obras por parte de un mismo autor, pero bajo nombres diferentes, afecte en modo alguno los indicadores de Delta y oculte la huella del autor en lo que respecta a este método: más aún, el principal objetivo de la atribución es crear una

2 Ver Schöch; Jannidis y Lauer; Calvo Tello et al.

3 Ver Reeve; Rebora y Salgaro; Stamou.

4 Ver Rybicki “Translation”; Rybicki y Heydel.

5 Ver Chojiński y Rybicki.

6 Ver Rybicki “Burrowing”; Hoover.

tecnología capaz de identificar al autor que se oculta bajo diferentes nombres. Como indican Evert y coautores, “el supuesto fundamental en la atribución de autoría es que los individuos tienen hábitos idiosincrásicos de uso del lenguaje, lo que lleva a una similitud estilística de los textos escritos por la misma persona” (254). Por lo tanto, esperamos que los seudónimos no creen una señal estilística fuerte. Solo un trabajo del que tenemos constancia muestra cómo los textos escritos bajo diferentes seudónimos muestran una diferencia estilística significativa, sin embargo no es un caso totalmente pertinente, pues no podemos decir si la diferencia se debe a la señal autoral o al género.⁷

El presente artículo está dedicado a uno de estos casos especiales en el ámbito de las literaturas romances, el de un poeta de lengua portuguesa que logró “engañar” una herramienta estilométrica bastante sólida. La atención se centra en este éxito literario invisible a los ojos normales (no digitales). Por desgracia, no podemos dedicar este texto a una descripción detallada de la metodología estilométrica de Delta. Remitimos el lector, para los detalles técnicos, a los numerosos artículos de Burrows, Rybicki y otros mencionados en la introducción.

También debemos mencionar nuestro artículo sobre material similar, publicado en inglés en la revista *Digital Scholarship in the Humanities* (Skorinkin y Orekhov). Allí se dan otros paralelos a nuestro estudio y se incluyen más detalles técnicos.

Fernando Pessoa y sus heterónimos

Se tratará el caso del célebre poeta portugués Fernando Pessoa (1888-1935), famoso por sus decenas de heterónimos. Los heterónimos se diferencian de los seudónimos en la medida en que detrás de ellos hay individuos totalmente contruidos, con sus propias biografías, descripciones del aspecto físico y otras formas de representar la individualidad. Este fenómeno no debe considerarse una desviación psíquica. Como escribió una vez G. H.

7 Ver Ujma. Este estudio trata sobre la creadora de novelas rosa, Nora Roberts, que escribió una novela policíaca futurista bajo el seudónimo de J. D. Robb. Los géneros de los textos creados bajo su nombre real y su seudónimo difieren notablemente, lo que puede tener también un impacto significativo en los indicadores estilométricos.

Mead, “una personalidad múltiple es en cierto sentido normal [...] Suele haber una organización de todo el yo con referencia a la comunidad a la que pertenecemos y a la situación en la que nos encontramos”. Y esto tampoco se limita a la interacción social: “Hay partes del yo que sólo existen para el yo en relación consigo mismo” (Mead y Schubert 142).

En sus textos explicativos, más tarde reunidos en *Páginas de Doutrina Estética*, Pessoa hace una distinción fundamental entre seudónimos y heterónimos (citado en Jones 254). Insiste en que un heterónimo no es solo un nombre, sino precisamente una identidad distinta capaz de expresar lo que el propio autor no puede expresar (Jones). Entre las 136 identidades poéticas creadas por Pessoa (Pessoa, *Eu sou*), tres se consideran los más importantes: Alberto Caeiro, Álvaro de Campos y Ricardo Reis. Los tres aparecieron en 1914. El propio Pessoa subrayaba constantemente la diferencia entre ellos en la forma de percibir el mundo y, en consecuencia, en la estilística de sus textos. Como demostraremos más adelante, esta diferencia puede resultar más fuerte que la unidad de la señal autoral. Los heterónimos no solo difieren significativamente entre sí y del propio Pessoa. Incluso podemos decir que, en la dimensión estilística, se comportan como autores distintos, es decir, difieren de Pessoa no menos que de otros autores reales que escribieron en portugués.

Existe una amplísima bibliografía sobre los heterónimos de Pessoa, en la que se analizan detalladamente los aspectos psicológicos, biográficos y poéticos del fenómeno. Nos limitamos a revisar trabajos de carácter panorámico, pues la mayoría de estudios sobre el tema se centran en aspectos ajenos a nuestra investigación.

Quizá señalar a Pessoa como un *unicum* capaz de confundir los métodos modernos de análisis estilométrico no se ajuste del todo a la realidad. Es más probable que debamos hablar de una especie de continuo en el que en un polo está un seudónimo, cuyos textos firmados no difieren de los del propio autor y, en el otro polo, una personalidad literaria completamente independiente, como las creadas por Pessoa. Sin embargo, la confirmación de esta hipótesis requiere más investigación.

El problema de la delimitación cuantitativa del estilo de Pessoa y sus heterónimos lo planteamos ya en 2017 (Orekhov), pero entonces no se

resolvió del todo debido al uso de métodos poco fiables de análisis de datos estilométricos. En ese trabajo se utilizó el método k-medias para agrupar los textos por palabras. En ninguna parte se había demostrado que este método fuera realmente fiable para indicar la autoría. Además, el corpus de textos utilizado era pequeño, sensible a la temática y, en general, no permitía llegar a resultados fiables. Ahora volvemos sobre este problema con una caja de herramientas diferente, con métodos que se han convertido en el *state-of-the-art* en este campo en las últimas dos décadas.

El análisis estilométrico en nuestro trabajo se realiza mediante la función *stylo* del paquete Stylo para el lenguaje R (Eder et al.). Realizamos tres series de experimentos estilométricos. En el primero, el corpus de investigación estaba formado por textos poéticos creados por el propio Pessoa y firmados o con su nombre (“*orthonymic Pessoa*”; Jones 254) o con el nombre de sus tres principales heterónimos. La fuente de textos fue una edición contemporánea de Pessoa.⁸ Hemos utilizado las ediciones digitalizadas disponibles, pues incluso suponiendo un trabajo textual no del todo exacto en la preparación de estos libros, se sabe que el método Delta no es sensible a diferencias y cambios puntuales. Un experimento en el que se compararon los resultados a partir de una edición textualmente correcta de *El Don apacible*, de Mijaíl Shólojov, y una versión “masiva” de esta novela que circula por Internet demostró que, desde el punto de vista del Delta, no hay diferencia entre estos textos (Velikanova y Orekhov).

Aquí investigamos hasta qué punto la estilística puede distinguir entre los trabajos firmados por los heterónimos y por el ortónimo de Pessoa. En la segunda serie, añadimos poetas contemporáneos de Pessoa al corpus de investigación. El objetivo principal de estos experimentos era comprobar cómo la diferencia estilística de los heterónimos se relaciona con la diferencia entre los distintos autores. En la tercera serie realizamos algunas pruebas adicionales. Los datos completos de los experimentos pueden consultarse en nuestro repositorio en línea (“*hackingstylometry*”).

8 Ver Pessoa, *Poesia*; Pessoa y de Campos; Pessoa y Caeiro; Pessoa y Reis.

Experimento 1: Pessoa vs. Pessoa

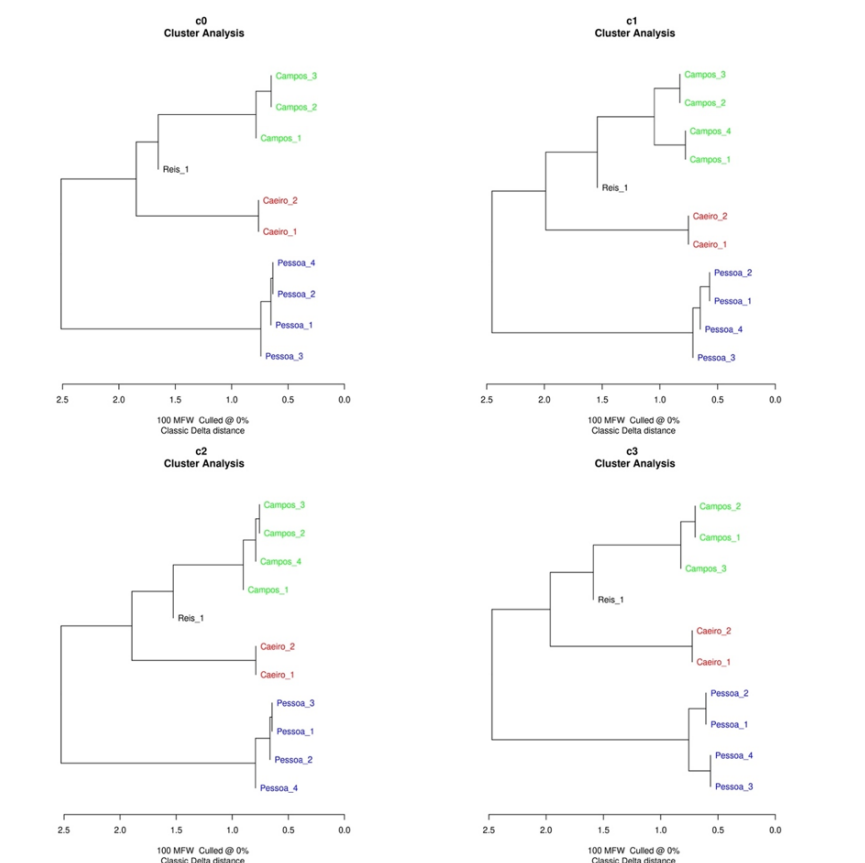
Para este experimento recopilamos un corpus de poemas de Pessoa firmados por uno de cuatro nombres: Fernando Pessoa, Ricardo Reis, Alberto Caeiro o Alvaro de Campos. Atribuir un texto concreto a uno de los heterónimos de Pessoa puede resultar no tan simple. Nos guiamos por la distribución de textos que figuraba en las ediciones electrónicas que utilizamos. Es posible, en este caso, que algunas de las decisiones editoriales sean erróneas. Pero los resultados que hemos obtenido sugieren que, en general, cada subcorpus de textos correspondiente a uno de los heterónimos resulta ser estilísticamente homogéneo. Tratamos cada uno de estos nombres como un autor independiente.

Dado que la estilometría requiere un gran conjunto textual para obtener resultados fiables, no trabajamos con poemas individuales, sino con un conjunto de poemas lo suficientemente grande como para contener al menos diez mil palabras. Así, todos los textos atribuidos a un mismo heterónimo u ortónimo fueron mezclados y divididos en *chunks*. Fernando Pessoa en cuanto tal fue representado por cinco ficheros, Alberto Caeiro por tres, Ricardo Reis por dos, los textos de Alvaro de Campos fueron divididos en tres o cuatro ficheros, dependiendo de cómo se distribuyeran las obras largas dentro del mismo *chunk*. Los poemas concretos no se dividieron y se colocaron íntegramente en el *chunk* al que fueron asignados al azar.

Creamos a propósito varias versiones de ese corpus para ver en qué medida el resultado dependía de la aleatoriedad de la distribución de los textos en los *chunks*. A continuación, aplicamos un procedimiento estilométrico estándar de agrupación jerárquica de los resultados basado en las cien palabras más frecuentes. La imagen 1⁹ muestra los resultados de los cuatro experimentos realizados cada vez con una nueva versión del corpus y una nueva distribución de los poemas en los *chunks*.

Como muestra la imagen 1, los resultados son estables y coherentes, es decir, no dependen de cómo se distribuyen los poemas individuales entre los distintos *chunks*. Delta reconoce con seguridad la diferencia entre el ortónimo Pessoa y los tres heterónimos, dado que Ricardo Reis, Alberto Caeiro y Álvaro de Campos muestran sus “voces” individuales.

9 Todas las tablas e imágenes son nuestras.

Imagen 1. Dendrogramas de *cluster* de la similitud estilométrica entre Pessoa y sus heterónimos

En aras de la pureza del experimento, es esencial que este corpus sea bastante homogéneo en cuanto a diversidad de géneros. A diferencia de otros autores que produjeron obras de otros géneros bajo seudónimos (por ejemplo, Mary Westmacott de Agatha Christie o Robert Galbraith de J. K. Rowling), las obras poéticas de Pessoa, firmadas con heterónimos, comparten más o menos la misma especificidad de género.

Así, podemos hablar de la “individualización” de los heterónimos. Sin embargo, el siguiente paso es averiguar hasta qué punto será estable la clasificación tras la introducción de otros autores en el campo de estudio.

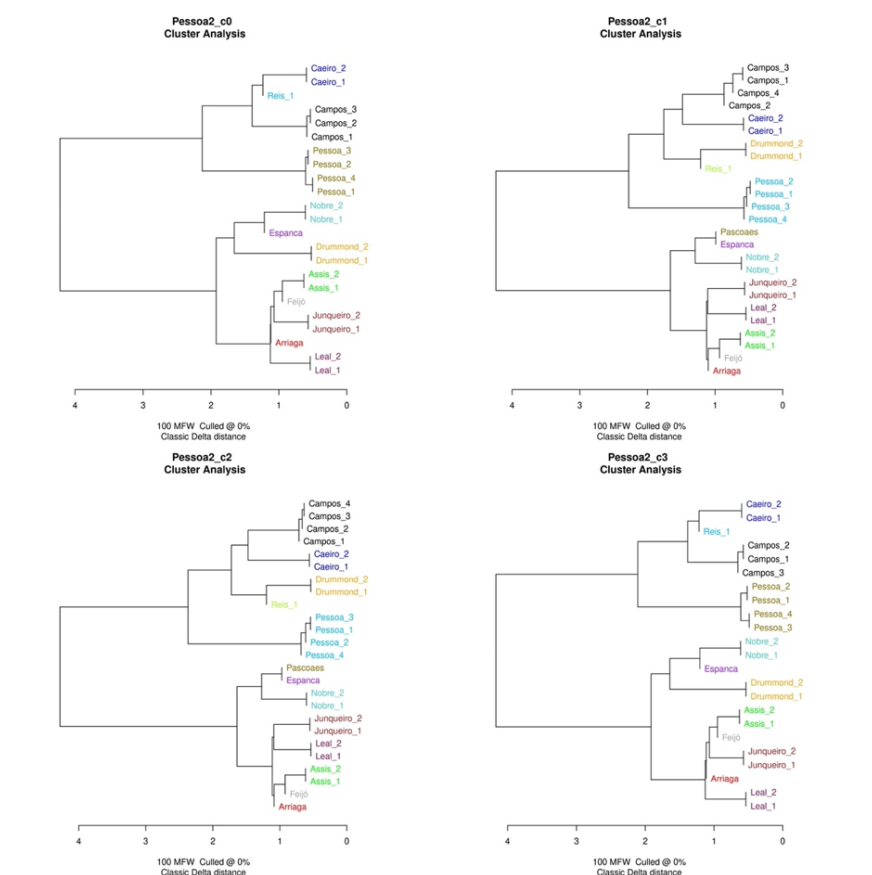
Experimento 2: Pessoa vs. otros poetas

En este punto, hemos añadido a nuestro corpus de investigación otros autores portugueses contemporáneos de Pessoa. Juola afirma que la cuestión de cuántos autores individuales se necesitan para experimentos estilométricos de este tipo sigue abierta, pero en general el orden de tres a siete parece razonable (106). Dado que el número de versiones digitalizadas de los textos de los poetas de interés era limitado, incluimos no solo a poetas portugueses, sino también brasileños. También hemos incluido en el corpus obras de poetas de la generación anterior a Pessoa, pues pretendíamos que nuestro corpus fuera más representativo. Lamentablemente, nuestra elección de autores se vio limitada por el hecho de que la poesía contemporánea de Pessoa en portugués está poco digitalizada. Hay muy pocos textos disponibles en formato electrónico, y necesitábamos colecciones textuales lo suficientemente amplias (al menos diez mil palabras) para realizar una comparación cuantitativa.

He aquí la lista de poetas cuyos textos sirvieron para compararlos con los de Pessoa y sus heterónimos: Manuel de Arriaga (1840-1917), Machado de Assis (1839-1908, brasileño), Florbela Espanca (1894-1930), Carlos Drummond de Andrade (1902-1987, brasileño), António Feijó (1859-1917), Guerra Junqueiro (1850-1923), António Gomes Leal (1848-1921) y António Pereira Nobre (1867-1900).

Con este corpus realizamos experimentos idénticos a los descritos anteriormente. Los poemas de cada autor se recopilaron en fragmentos de al menos diez mil palabras cada uno. Creamos a propósito varias versiones del corpus, cada vez distribuyendo aleatoriamente los poemas en los *chunks*. De este modo, intentamos eliminar la aparición de un único efecto aleatorio. A continuación, aplicamos un procedimiento estilométrico estándar de agrupación jerárquica de los resultados basado en las cien palabras más frecuentes. La imagen 2 muestra los datos obtenidos a través de varias agrupaciones de los poemas.

Imagen 2. Dendrogramas de *cluster* de la similitud estilométrica entre Pessoa, sus heterónimos y una serie de poetas lusófonos contemporáneos de Pessoa (cada clusterización corresponde a una distribución aleatoria distinta de poemas en archivos)

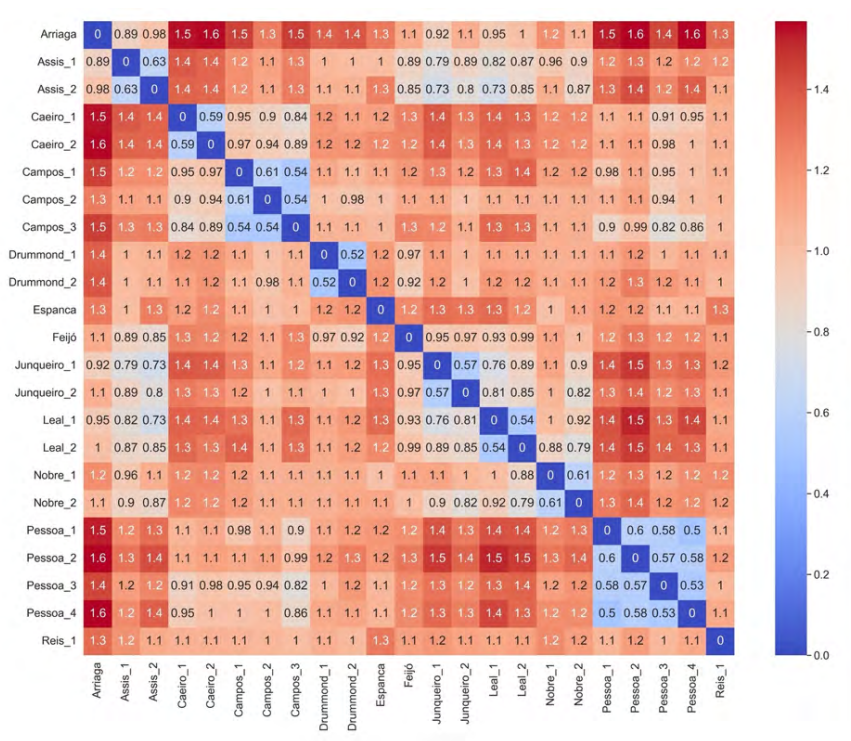


Aquí nos espera un descubrimiento clave: los heterónimos crean una señal particular y distinta a la del autor. De hecho, vemos que cada uno de los heterónimos de Pessoa está agrupado de tal manera que parece un autor independiente, lo que contradice la premisa básica de la estilometría. La incorporación de otros poetas al corpus no ha roto esta tendencia. Contrariamente a lo que cabría esperar, basándonos en nuestros conocimientos sobre el funcionamiento del método Delta, la estilometría se mostró incapaz de distinguir los textos escritos por el mismo autor con

un nombre diferente de los textos realmente creados por otros autores. El panorama no cambiaba incluso al elevar de cien a quinientas el umbral de las palabras más frecuentes utilizadas en el recuento. La “seudo” señal de autoría de los heterónimos permanece estable. Estos resultados también se verificaron mediante la prueba *classify()* y la precisión de la predicción volvió a ser del 100 %.

La voz particular de los heterónimos también se aprecia en el mapa de calor que hemos construido a partir de las distancias entre los textos calculadas con Delta (imagen 3).

Imagen 3. Visualización en mapa de calor de las distancias estilométricas entre Pessoa, sus heterónimos y una serie de poetas lusófonos contemporáneos de Pessoa

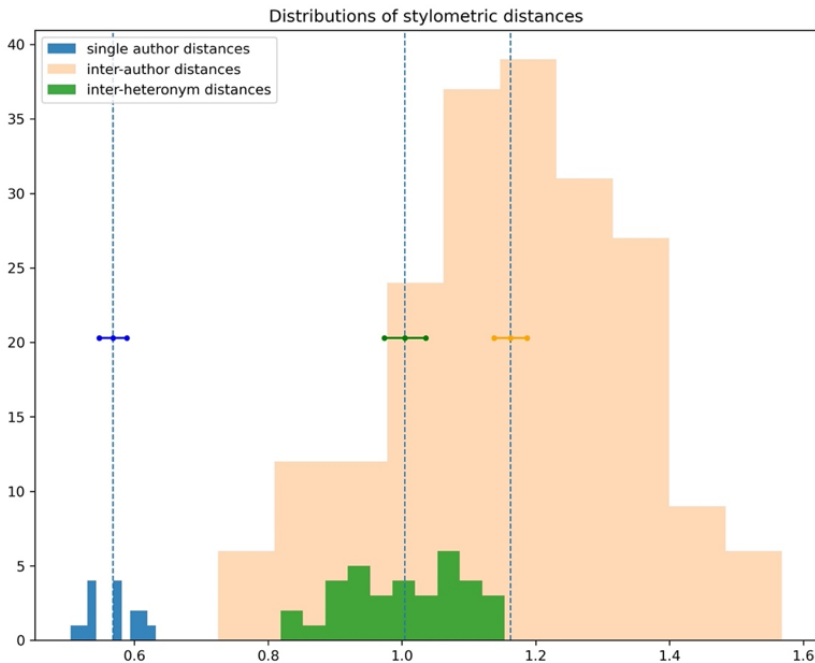


El color azul indica una mayor similitud estilística. Es claramente visible que cada autor muestra similitudes entre sus textos y divergencias con los textos de otros autores. Esta situación se fija para todos los heterónimos

(Caeiro, de Campos, Reis y el ortónimo Pessoa) del mismo modo que para otros “autores naturales” como Drummond o Machado de Assis.

La especificidad estilística de los heterónimos de Pessoa también puede apreciarse en la visualización que refleja las distancias para todos los pares de textos de nuestro experimento. El azul en la imagen 4 muestra las distancias entre textos del mismo autor (por ejemplo, entre Drummond_1 y Drummond_2 o Pessoa_1 y Pessoa_3), excluyendo las distancias entre los textos de Pessoa y sus heterónimos. El beige representa las distancias entre textos de diferentes autores (por ejemplo, entre Drummond_1 y Pessoa_3 o Assis_2 y Espanca). El verde representa las distancias entre los textos de los heterónimos y los textos del ortónimo Pessoa (por ejemplo, Pessoa_4 y Campos_2).

Imagen 4. Distribuciones de tres grupos de distancias estilométricas: las distancias de un solo autor, las distancias entre autores y las distancias entre heterónimos; las líneas verticales discontinuas son las medias de cada distribución, mientras que las líneas horizontales azules son intervalos de confianza del 95 %



Así, podemos ver claramente que las distancias entre los textos de Pessoa y sus heterónimos se aproximan más a los valores de las distancias entre autores “naturalmente” diferentes. En la situación habitual, la distancia entre los textos del mismo autor es notablemente menor que la distancia entre los textos de Pessoa, Reis, de Campos y Caeiro. El hecho de que los *clusters* de Caeiro, de Campos, Reis y Pessoa destaquen sistemáticamente tanto en el primer experimento como en el segundo indica la estabilidad del experimento estilométrico. Como vemos, en este caso el cambio de corpus no afecta al resultado, lo que demuestra la fuerza de la señal autoral de los heterónimos.

Experimento 3: Un heterónimo

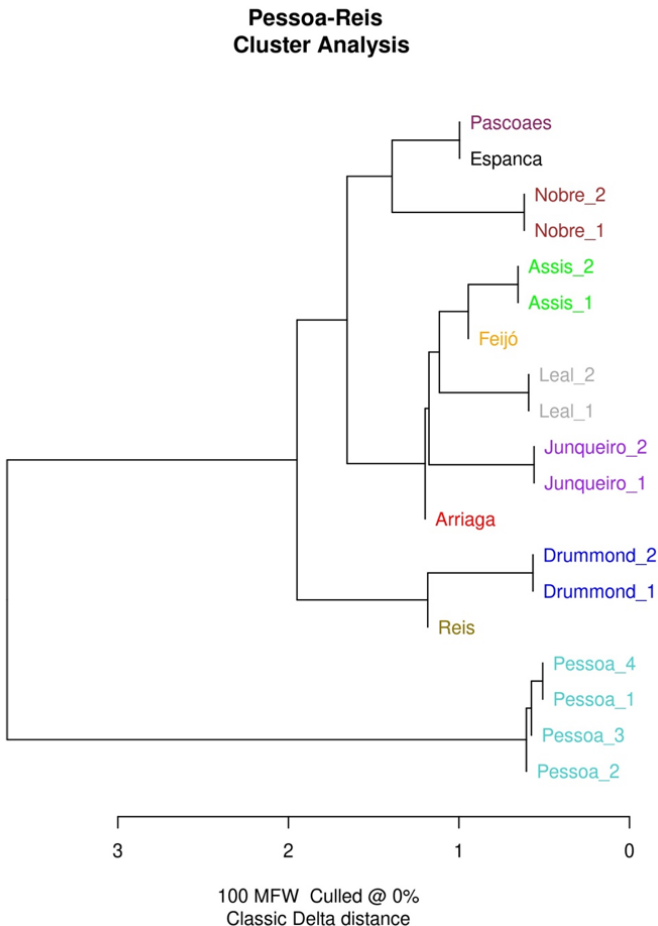
Es bien sabido que para los experimentos estilométricos es necesario disponer de al menos dos archivos con textos de cada autor. Si solo hay un texto de la longitud adecuada para un autor en el corpus de investigación, puede agruparse junto con cualquier otro autor. Un ejemplo de esto se puede ver en la imagen 2, en la cual vemos solamente un archivo con los poemas de Florbela Espanca. Lamentablemente, no hay muchas obras de esta poeta disponibles en formato electrónico, por lo que solo hemos podido incluir un conjunto de poemas en el corpus. La figura muestra que este único archivo se comporta de forma incoherente, agrupándose con diferentes autores, de un experimento a otro, en función de la configuración dada. Se trata de limitaciones naturales de la estilometría: cuantos más archivos de cada autor se incluyan en el corpus de investigación, más estables y fiables serán los resultados.

En nuestro estudio de los heterónimos, sin embargo, tiene sentido intentar intencionadamente que los resultados sean menos fiables. Intentaremos utilizar una configuración de corpus en la que solo haya un archivo con los textos de un heterónimo. En este caso, no podrá formar un cluster separado como en experimentos anteriores, pero ¿se agrupará junto con los textos del propio Pessoa? Como se verá más adelante, a veces es así. No obstante, la proximidad estilística de los heterónimos de Pessoa sigue siendo menos pronunciada que la de los textos del mismo autor. Así, la singularidad de los heterónimos sigue siendo visible.

Pessoa vs. Reis

En nuestro primer intento de confundir la señal de un heterónimo, solo incluimos un archivo con textos de Ricardo Reis, y no incluimos textos de Campos y Caeiro. En el mismo corpus estaban todos los demás poetas y el ortónimo Pessoa. Los resultados se muestran en la imagen 5.

Imagen 5. Dendrogramas de cluster de similitud estilométrica entre una única muestra de Reis, Pessoa, y una serie de poetas lusófonos contemporáneos de Pessoa (cada clusterización corresponde a una distribución aleatoria distinta de poemas en archivos)



Los textos de Reis según estos resultados no se fusionan con Pessoa. Sin embargo, en este caso el algoritmo de agrupación por defecto Stylo no ofrece una imagen muy clara. Cuando comprobamos los valores de las distancias entre textos calculados por el método Delta, resulta que los más próximos a Reis son los textos de Pessoa, a pesar de que en el diagrama Reis está en el mismo cluster que Drummond.

La tabla 1 muestra las distancias entre Reis y los cinco fragmentos más cercanos a él en términos de Delta. Los valores se ordenan de mayor a menor distancia.

Tabla 1.

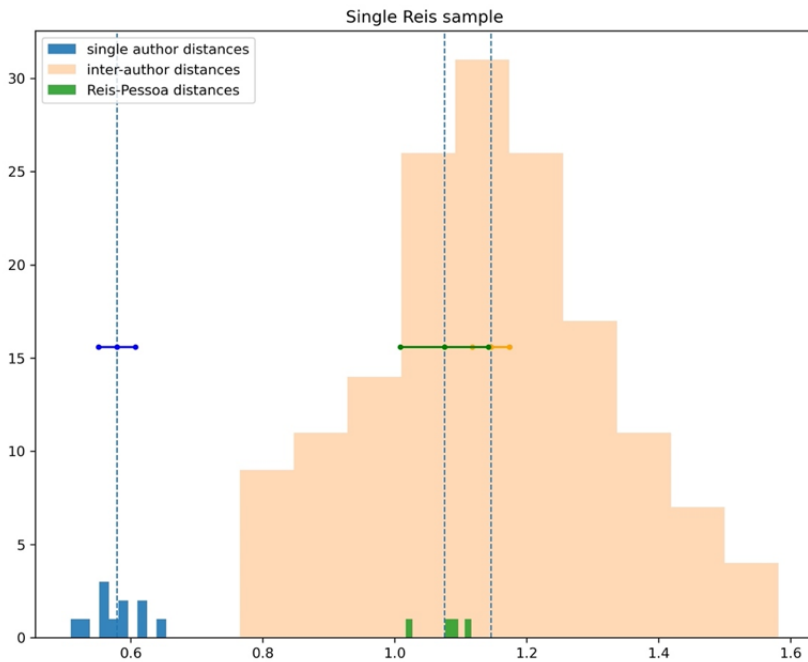
Nombre del archivo	Distancia estilométrica de los textos de Reis
Drummond_2	1,016524
Pessoa_3	1,016714
Drummond_1	1,040993
Pessoa_1	1,078667
Pessoa_4	1,089540

Gracias a la tabla podemos ver que los *chunks* con textos de Pessoa son los más cercanos a Drummond_2, pero que, en general, las distancias de Reis a Pessoa y a Drummond no son radicalmente diferentes. También resulta interesante que los textos de Reis se agruparan con los de un poeta brasileño y no con uno portugués. Este hecho llama la atención gracias a un dato biográfico de Reis, inventado por Pessoa: Ricardo Reis, tal y como lo concibió Pessoa, se marchó de Portugal a Brasil, y por lo cual es el único que ha “sobrevivido” tanto a los otros heterónimos (para cada uno de ellos Pessoa describió el fallecimiento) como a su creador. El final de la vida de Reis lo imagina otro autor, José Saramago, en su novela *El último año de la vida de Ricardo Reis* (1984). Así, por una divertida coincidencia, el diagrama estilométrico capta en este caso una especie de señal “brasileña”.

En comparación, la distancia entre Pessoa_3 y Pessoa_4 es dos veces menor: 0,51. En general, las distancias entre textos del mismo autor suelen situarse entre 0,5 y 0,65, mientras que las distancias entre textos de autores diferentes suelen ser superiores a 1 (ver tabla 1). A continuación, en la imagen

6, se muestran dos distribuciones: el histograma azul muestra todas las distancias entre textos de distintos autores y el histograma naranja muestra las distancias entre textos de un mismo autor (por ejemplo, Drummond_1 y Drummond_2). Las distancias entre Reis y Pessoa se muestran en verde.

Imagen 6. Distribuciones de tres grupos de distancias estilométricas: las distancias de un solo autor, las distancias entre autores y las distancias entre Reis y Pessoa; las líneas verticales discontinuas son las medias de cada distribución, mientras que las líneas horizontales azules son intervalos de confianza del 95 %

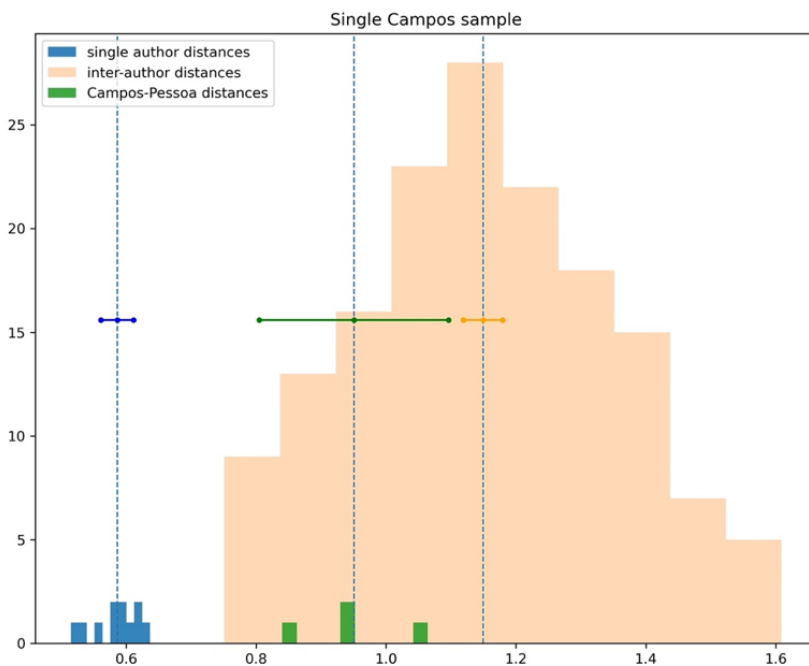


El histograma muestra de forma convincente que la distancia estilométrica entre Pessoa y Reis se encuentra en la misma zona en la que suelen encontrarse las distancias entre textos de autores diferentes (físicamente diferentes). Se repiten así los resultados del experimento anterior, y se confirma de nuevo la tesis de que Pessoa ha logrado crear heterónimos estilísticamente individuales.

Pessoa vs de Campos y Pessoa vs Caeiro

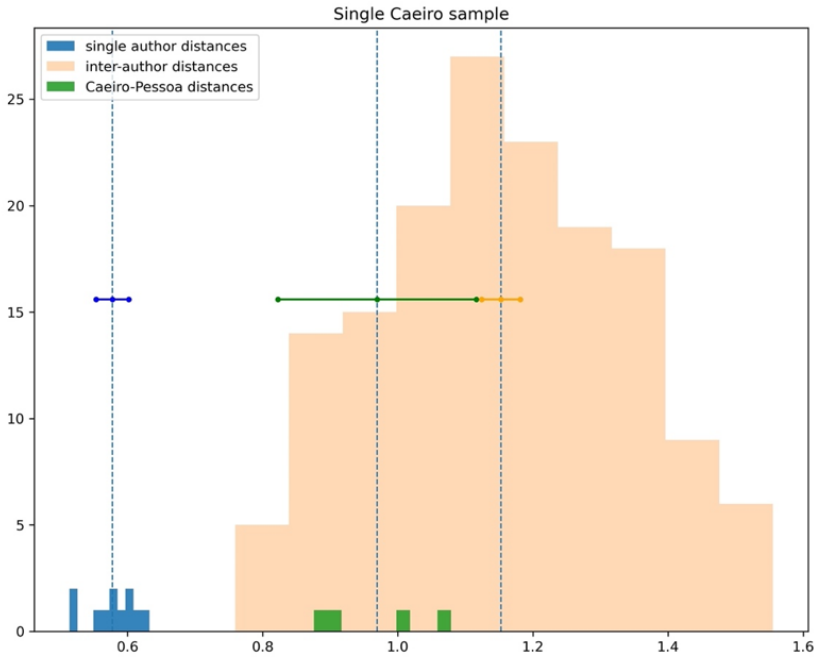
Los otros dos heterónimos del experimento descrito muestran resultados similares. Los textos sueltos de Caeiro y de Campos tienen la menor distancia estilística exactamente con Pessoa, pero al mismo tiempo el propio valor de la distancia entre los textos de los heterónimos y Pessoa es mucho mayor que entre los textos de un solo autor. Estos datos se ilustran en las imágenes 7 y 8.

Imagen 7. Distribuciones de tres grupos de distancias estilométricas: las distancias de autor único, las distancias entre autores y las distancias entre Campos y Pessoa



Estos resultados son plenamente coherentes con todo lo descrito anteriormente: los heterónimos creados por Pessoa demuestran lo que podría denominarse autonomía estilística, característica de individuos y no de personalidades ficticias.

Imagen 8. Distribuciones de tres grupos de distancias estilométricas: las distancias de autor único, las distancias entre autores y las distancias entre Caeiro y Pessoa



Conclusiones

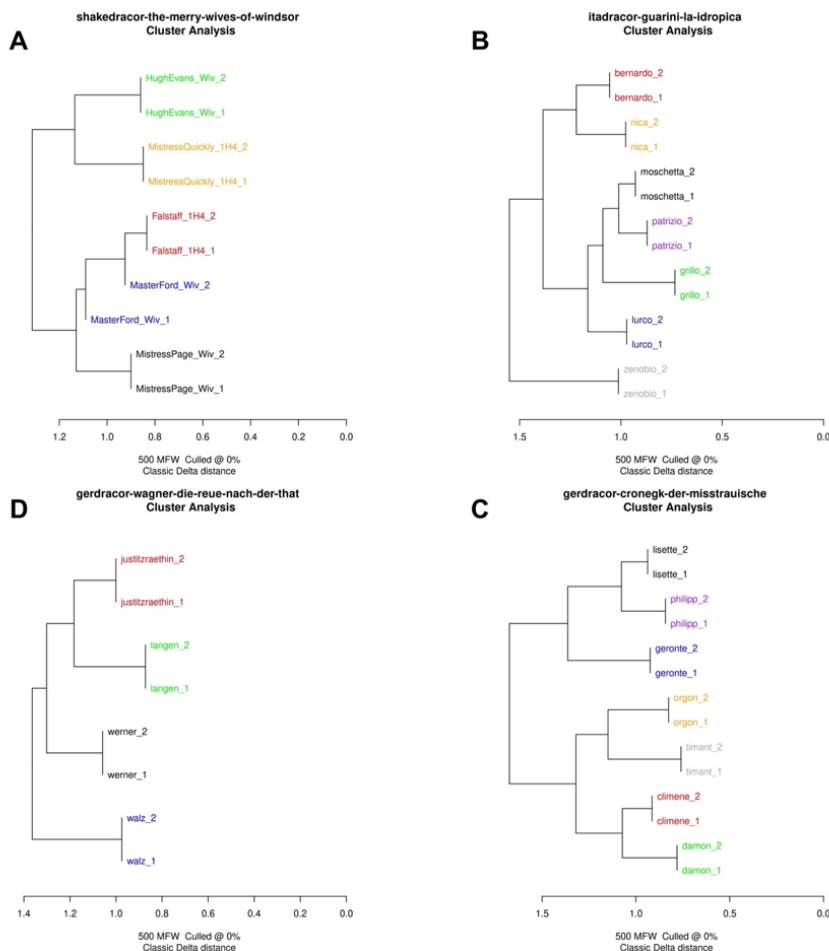
No podemos decir que la capacidad del autor para crear “voces” estilométricamente discernibles sea algo totalmente inesperado. Como mencionamos en la reseña de la introducción, en principio se sabe desde hace tiempo que los personajes de ficción en algunos casos tienen sus propias “voces” construidas para ellos intencionadamente.

También se sabe que Pessoa consideraba sus heterónimos como algo parecido a los personajes de un drama (Jones 255). Esta analogía aparece a menudo en las cartas de Pessoa:

Lo que yo llamo literatura insincera no es aquella análoga a la de Alberto Caeiro, Ricardo Reis o de Álvaro de Campos [...]. Esta se siente en la persona de otro; está escrita dramáticamente, pero es sincera (en mi sentido serio de

la palabra) como lo es lo que dice el Rey Lear, que no es Shakespeare, sino una creación suya.¹⁰ (Pessoa, “Carta” s. p.)

Imagen 9. Dendrogramas de cluster de similitud estilométrica entre las voces de los personajes de cuatro obras de teatro



¹⁰ La traducción es nuestra. En original: “O que eu chamo literatura insincera não é aquela análoga à do Alberto Caeiro, do Ricardo Reis ou do Álvaro de Campos [...]. Isso é sentido na pessoa de outro; é escrito dramaticamente, mas é sincero (no meu grave sentido da palavra) como é sincero o que diz o Rei Lear, que não é Shakespeare, mas uma criação dele.”

Esta visión de sus heterónimos por parte de Pessoa implica precisamente el deseo de crear un estilo específico para ellos. Nuestros experimentos demuestran que Pessoa ha cumplido esta tarea.

Para comparar el estatus estilístico de los personajes dramáticos con heterónimos, llevamos a cabo una serie de experimentos con textos dramáticos almacenados por el proyecto Dracor (Fischer et al.). La imagen 9 muestra algunos resultados representativos (en orden cronológico): a) *The Merry Wives of Windsor* de William Shakespeare, b) *La idropica* de Giovanni Battista Guarini, c) *Die Reue nach der That* de Heinrich Leopold Wagner, d) *Der Misstrauische* de Johann Friedrich von Cronegk.

El conjunto de discursos de un personaje de cada obra se dividió aleatoriamente en dos fragmentos y, a continuación, se realizaron las operaciones estilométricas estándar descritas anteriormente para los textos de Pessoa, sus heterónimos y sus contemporáneos. Los resultados que vemos a través del estudio de los textos dramáticos muestran que muchos autores de obras de teatro son capaces de crear personajes cuyas voces son marcadamente diferentes desde la óptica estilométrica.

Así, desde el punto de vista estilométrico, se puede describir la vida literaria de Fernando Pessoa como un acto dramático con un cambio constante de papeles en el escenario. Probablemente, Pessoa no fue el único capaz de esta construcción estilística de voces “autorales” independientes. Lo sorprendente y aparentemente especial de Pessoa es que sus heterónimos no se crearon para ser usados una sola vez: hay cientos de poemas de Caetano de Campos y Reis, por no mencionar el ortónimo de Pessoa. Estas obras no siguen una única forma reconocible, ni están subordinadas a un tema fácil de identificar. Y, al mismo tiempo, su singularidad estilística es tan claramente distinguible a través de la estilística que, sin conocer la situación, podríamos pensar que se trata en realidad de personas diferentes del mundo real que escribieron estos poemas. Esto nos lleva de nuevo a la conclusión de que los rasgos estilométricos no son una verdadera “huella dactilar”. La analogía forense en este caso no es más que un truco “publicitario” que debe tomarse con cautela.

Obras citadas

- Burrows, John. “Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship”. *Literary and Linguistic Computing*, vol. 17, núm. 3, 2002, págs. 267-287. DOI: <https://doi.org/10.1093/lc/17.3.267>
- Burrows, John y Hugh Craig. “Authors and Characters”. *English Studies*, vol. 93, núm. 3, 2012, págs. 292-309. DOI: <https://doi.org/10.1080/0013838X.2012.668786>.
- Calvo Tello, José, et al. “Neutralising the Authorial Signal in Delta by Penalization: Stylometric Clustering of Genre in Spanish Novels”. *Digital Humanities 2017 Book of Abstracts*. Montréal, Canadá. Montréal, McGill University/Université de Montréal, 2017, págs. 181-183.
- Choiński, Michał y Jan Rybicki. “Jonathan Edwards and Thomas Foxcroft : Pursuing Stylometric Traces of the Editor”. *Amerikastudien*, vol. 63, núm. 2, 2018, págs. 141-158.
- Eder, Maciej y Jan Rybicki. “Go Set a Watchman while we Kill the Mockingbird in Cold Blood, with Cats and Other People”. *11th Annual International Conference of the Alliance of Digital Humanities Organizations, DH 2016, Krakow, Poland, July 11-16, 2016, Conference Abstracts*. Editado por Maciej Eder y Jan Rybicki. Alliance of Digital Humanities Organizations (ADHO), 2016, págs. 184-186.
- Eder, Maciej, Jan Rybicki y Mike Kestermont. “Stylometry with R: a package for computational text analysis”. *R Journal*, vol. 8, núm. 1, 2016, págs. 107-121. DOI: <https://doi.org/10.32614/rj-2016-007>
- . “Visualization in stylometry: Cluster analysis using networks”. *Digital Scholarship in the Humanities*, vol. 32, núm. 1, 2017, págs. 50-64. DOI: <https://doi.org/10.1093/lc/fqv061>
- Evert, Stefan, et al. “Understanding and explaining Delta measures for authorship attribution”. *Digital Scholarship in the Humanities*, vol. 32, núm. suppl. 2, 2017, págs. ii4-iii16. DOI: <https://doi.org/10.1093/lc/fqx023>
- Fischer, Frank, et al. “Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama”. *Proceedings of DH2019: “Complexities”*, Utrecht University, 2019. DOI: <http://doi.org/10.5281/zenodo.4284002>.

- Forsyth, Richard y David Holmes. "Feature finding for text classification". *Literary and Linguistic Computing*, vol. 11, 1996, págs. 163-74. DOI: <https://doi.org/10.1093/lc/11.4.163>
- García, Antonio y Javier Calle-Martín. "Function Words in Authorship Attribution Studies". *Literary and Linguistic Computing*, vol. 22, 2006, págs. 49-66. DOI: <https://doi.org/10.1093/lc/fql048>
- "hackingstylometry". Zenodo.org. 31 de mayo de 2023. Web. 1 de junio de 2023. DOI: <https://doi.org/10.5281/zenodo.7992417>
- Hoover, David L. "The microanalysis of style variation". *Digital Scholarship in the Humanities*, vol. 32, núm. suppl. 2, 2017, págs. ii17-ii30. DOI: <https://doi.org/10.1093/lc/fqx022>
- Jannidis, Fotis y Gerhard Lauer. "Burrows's Delta and Its Use in German Literary History". *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*. Editado por Matt Erlin y Lynne Tatlock. Martlesham, Camden House, 2014, págs. 29-54.
- Jones, Marilyn Scarantino. "Pessoa's Poetic Coterie: Three Heteronyms and an Orthonym". *Luso-Brazilian Review*, vol. 14, núm. 2, 1977, págs. 254-262.
- Juola, Patrick. "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions". *Digital Scholarship in the Humanities*, vol. 30, núm. suppl. 1, 2015, págs. i100-i113. DOI: <https://doi.org/10.1093/lc/fqv040>
- Lustig, Bette H. "Emile Ajar Demystified". *The French Review*, vol. 57, núm. 2, 1983, págs. 203-12.
- Mead, George Herbert y Cornelius Schubert. *Mind, self and society*, vol. 111, Chicago, University of Chicago Press, 1934.
- Mosteller, Frederick y David L. Wallace. "Inference in an Authorship Problem". *Journal of the American Statistical Association*, vol. 58, núm. 302, 1963, págs. 275-309. DOI: <https://doi.org/10.2307/2283270>
- Orekhov, Boris. "Klasterizacija geteronimov Pessoa: mozhet li komp'juter otlichit' de Kampusha ot Rejsa?" *Podarok Natalii Azarovoj: Sbornik statej k jubileju N. M. Azarovoj*. Moscú, Kul'turnaja revoljucija, 2017, págs. 41-42.
- Pessoa, Fernando. "Carta a Armando Côrtes-Rodrigues – 19 Jan. 1915. *Arquivo Pessoa – Obra Édita*. Web. 16 de octubre de 2022.
- . *Eu sou uma antologia: 136 autores fictícios*. Lisboa, Tinta-da-china, 2016.
- . *Poesia Completa Ortônima de Fernando Pessoa*. Nostrum Editora, 2014.
- Pessoa, Fernando y Alberto Caeiro. *Poesia Completa de Alberto Caeiro*. Nostrum Editora, 2013.

Pessoa, Fernando y Álvaro de Campos. *Poesia Completa de Álvaro de Campos*. Nostrum Editora, 2014.

Pessoa, Fernando y Ricardo Reis. *Poesia Completa de Ricardo Reis*. Nostrum Editora, 2013.

Rebora, Simone y Massimo Salgaro. “Is ‘Late Style’ Measurable? A Stylometric Analysis of Johann Wolfgang Goethe’s, Robert Musil’s, and Franz Kafka’s Late Works”. *Elephant&Castle*, núm. 18, 2018, págs. 4-39.

Reeve, Jonathan Pearce. “Does ‘Late Style’ Exist? New Stylometric Approaches to Variation in Single-Author Corpora”. *Digital Humanities 2018 Book of Abstracts*, DH 2018, junio 26, 27, 28 y 29 de 2018, CDMX, México. México, El Colegio de México, UNAM y RedHD, 2018, págs. 478-480.

Rybicki, Jan. “Burrowing into Translation: Character Idiolects in Henryk Sienkiewicz’s Trilogy and its Two English Translations”. *Literary and Linguistic Computing*, vol. 21, núm. 1, 2006, págs. 91-103. DOI: <https://doi.org/10.1093/lc/fqh051>

———. “Translation and delta revisited: When we read translations, is it the author or the translator that we really read?”. *Digital Humanities 2009*, DH 2009, junio 20, 21, 22, 23, 24 y 25 de 2009, College Park, Estados Unidos. College Park, University of Maryland, 2009, págs. 245-247.

Rybicki, Jan y Magda Heydel. “The stylistics and stylometry of collaborative translation: Woolf’s Night and Day in Polish”. *Literary and Linguistic Computing*, vol. 28, núm. 4, 2013, págs. 708-717.

Sansone, David. “Stylistic Characterization in Plato: Nicias, Alcibiades, and Laches”. *Greek Roman and Byzantine Studies*, vol. 58, 2018, págs. 156-176.

Schöch, Christof. “Fine-tuning Stylometric Tools: Investigating Authorship and Genre in French Classical Theater”. *Digital Humanities 2013 Conference Abstracts*, DH 2013, julio 16, 17, 18 y 19 de 2013, Lincoln, Estados Unidos. Lincoln, Universidad de Nebraska, 2013, págs. 383-385.

Skorinkin, Daniil. “Semantic markup of literary texts for quantitative scholarship in philology (on the basis of Leo Tolstoy’s War and peace)”. Tesis doctoral, Universidad Nacional de Investigación “Escuela Superior de Economía”, 2019.

Skorinkin, Daniil y Boris Orekhov. “Hacking stylometry with multiple voices: Imaginary writers can override authorial signal in Delta”. *Digital Scholarship in the Humanities*, 2023. DOI: <https://doi.org/10.1093/lc/fqad012>

- Stamou, Constantina. “Stylochronometry: Stylistic Development, Sequence of Composition, and Relative Dating”. *Literary and Linguistic Computing*, vol. 23, núm. 2, junio de 2008, págs. 181-199. DOI: <https://doi.org/10.1093/llc/fqm029>
- Ujma, Alicja. “One Author, Two Pen Names. A Stylometric Analysis of Novels by Nora Roberts/J.D. Robb”. Tesis doctoral, Universidad Jaguelónica, 2019.
- van Dalen-Oskam, Karina y Joris van Zundert. “Delta for Middle Dutch-Author and Copyist Distinction in Walewein”. *Literary and Linguistic Computing*, vol. 22, núm. 3, 2007, págs. 345-362, DOI: <https://doi.org/10.1093/llc/fqm012>
- Velikanova, Natalian y Boris Orekhov. “Tsifrovaia tekstologia: atrobutsia teksta na primere romana M. A. Sholajova”. *Tiji Don. Mir Sholajova*, vol. 1, núm. 11, 2019, págs. 70-83.

Sobre los autores

Daniil Skorinkin es postdoctorando y coordinador de Humanidades Digitales en la Universidad de Potsdam. Anteriormente fue profesor asociado en la Facultad de Humanidades de la Escuela Superior de Economía de Moscú. Su doctorado en lingüística y estudios literarios se centró en la aplicación de métodos digitales al análisis de la literatura rusa. Sus áreas de investigación incluyen la estilometría, el análisis de redes y el uso de herramientas de procesamiento de lenguaje natural (PNL) para la investigación en humanidades. Imparte clases de Humanidades Digitales, Estudios Literarios Computacionales y Python para Humanidades.

Sobre el artículo

Este artículo inédito fue traducido del ruso por Anastasia Belousova (Universidad Nacional de Colombia). Hace parte del proyecto de investigación “La traducción poética múltiple en Rusia: un enfoque lingüístico-estadístico y cuantitativo”, que se desarrolla en el Universidad Nacional de Investigación “Escuela Superior de Economía” con el apoyo del Fondo Científico Ruso (beca 23-28-01201; <https://rscf.ru/project/23-28-01201/>). Agradecemos a Frank Fischer, Olga Alieva, Jan Rybicki y Artjoms Šeļa, cuyos consejos ayudaron a mejorar notablemente este estudio. Al mismo tiempo, todos los errores y deficiencias quedan en la conciencia de los autores.

Boris Orekhov es profesor asociado de la Escuela Superior de Economía (HSE University) e investigador principal del Instituto de Literatura Rusa (Casa Pushkin) de la Academia Rusa de Ciencias. Sus intereses de investigación giran en torno a la versología, las humanidades digitales y la lingüística de corpus.