

escondidas (los *única*) y otras ya conocidas, ahora iluminadas con nueva luz gracias a sus peculiares variantes.

El *Cors625*, pues, con sus poemas únicos (16 en total), con sus glosas únicas (7) y el mérito de brindar “versiones distintas y aisladas del resto de la tradición” (p. 8); con sus textos de tradición múltiple (10) y paternidad ilustre (Lope de Vega, Quevedo), restituye la significativa variedad y vitalidad de la lírica española de tema amoroso, burlesco, jocoso y satírico, que en la primeras décadas del siglo XVII circulaba por las cortes italianas y, en varias ocasiones, exclusivamente en la romana. Al incluir géneros poéticos distintos: décimas, romances, villancicos, seguidillas, el *Cors625* “custodia dentro de sus textos y ritmos de guitarra el son de la música que sirvió a los protagonistas [del “Gran teatro del mundo romano”] para alegrar su ocio y, quizás aliviar sus penas, al fin y al cabo, para vivir” (93).

DOI 10.14672/1.2024.2485

**Francisco Javier Blasco Pascual;
Cristina Ruiz Urbón, *Análisis
de textos desde la estilometría,*
Salamanca, Universidad de
Salamanca, 2022, 216 pp. ISBN
9788413117638**

**Alejandro García-Reidy
IEMRhd, Universidad de Salamanca**

¿Quién teme a la estilometría? La consolidación y expansión de herramientas di-

gitales en las últimas décadas ha abierto nuevas posibilidades de investigación en humanidades, incluyendo el ámbito de los estudios literarios y lingüísticos hispánicos. Uno de los acercamientos más desarrollados, y que mayor utilidad actual y potencialidad futura muestra para abordar una serie de problemas y preguntas, es el de la estilometría, es decir, el análisis estadístico de textos desde una variedad de parámetros (frecuencia de palabras, longitud de oraciones, colocaciones o correlaciones de términos, etc.). Si bien el análisis de palabras está en el corazón de los estudios filológicos, es la confluencia de las posibilidades de cuantificación que permite un ordenador y el desarrollo de aplicaciones específicas lo que ha revolucionado el campo de la estilometría digital en fechas recientes y lo que le augura un futuro productivo. Este libro, a cargo de dos especialistas en el campo, busca ofrecer al investigador una introducción, panorámica y profunda a la vez, a la estilometría digital y a sus posibilidades aplicadas a textos en español, sobre todo literarios. La mayoría de los ejemplos de análisis que se ofrecen se basan en un amplio espectro de la literatura española, desde textos de los siglos XVI y XVII, pasando por Espronceda, Galdós o Pardo Bazán, hasta obras de Miguel Delibes, Muñoz Molina o Almudena Grandes, aunque también se usan algunos textos políticos.

Conviene dejar claro que este libro no es un manual con los pasos que hay que seguir para utilizar herramientas de análisis estilométrico. El investigador deberá buscar las instrucciones de uso en otra parte, preferentemente en la red, que es el espacio que, por su fluidez, permite ofrecer tutoriales actualizados. La buena noticia es que

no hace falta tener grandes conocimientos informáticos para adentrarse en la estilometría y aprovechar sus posibilidades, y el manual en red preparado por José Manuel Fradejas al que remiten los autores del presente volumen es un punto de partida al día y didáctico para el interesado. Este libro, en cambio, explica y ejemplifica cómo se está empleando la estilometría para analizar y entender características de textos, y muestra el potencial que permite la cuantificación de fenómenos lingüísticos en grandes cantidades de datos, es decir, de palabras; por ejemplo, para proponer la identificación de autoría de un texto; detectar cambios en el estilo de un autor que respondan a cuestiones de género literario o de cronología; cuantificar semejanzas y diferencias entre textos; o detectar patrones léxicos que permitan la identificación automática de temas, por mencionar algunos de ellos. Es, por tanto, un libro que invita al neófito (o al receloso) a descubrir cómo la estilometría puede ayudarle a plantear y contestar preguntas relacionadas con sus propias líneas de investigación.

El libro está dividido en siete capítulos, más unas conclusiones, bibliografía y apéndices finales. Si el primero de los capítulos ofrece una presentación básica de lo que es la estilometría, el segundo se centra en los principales protocolos que conviene seguir para preparar el corpus que se quiera analizar, con algunas referencias a corpus en español públicos y accesibles en red. El tercer capítulo empieza a desgranar cómo la cuantificación de un texto, incluso al nivel de caracteres, permite empezar a discriminar entre autores a partir de mínimos rasgos estilísticos.

Sin embargo, es el cuarto capítulo, de-

dicado al análisis léxico, el primero especialmente jugoso para el lector, pues ahí se presentan una serie de ejemplos que muestran claramente la utilidad de la estilometría para el análisis de textos literarios en español. Así, los casos que se desarrollan de las clasificaciones y agrupaciones a partir de las palabras más frecuentes de un corpus ilustran su función como señal autorial con una altísima tasa de fiabilidad, con recursos que permiten eliminar falsos positivos o detectar la presencia de varias manos en un texto pluriautorial. La estilometría léxica identifica asimismo las palabras más representativas de un texto en contraste con el conjunto del corpus. Esto permite detectar temas o rasgos de una obra particular en relación con textos semejantes, así como localizar patrones léxicos más profundos, propios del estilo de un autor, que emergen solo cuando se cuantifican. El análisis léxico también puede ser el punto de partida para acercamientos cualitativos que aprovechen la capacidad cuantificadora de la estilometría digital, desde la datación de textos atendiendo a cambios de estilo en el tiempo a trazar una evolución de las emociones negativas o positivas en un texto a partir de la asignación de una carga emocional a cada palabra. Ahora, bien, como señalan los autores del libro, los diccionarios y etiquetados automáticos todavía no están tan pulidos para el español como lo están para el inglés, de manera que algunos tipos de análisis todavía distan de la fineza que podrían presentar.

El capítulo quinto se centra en el análisis morfológico desde la estilometría, sobre todo en relación con el etiquetado automático de categorías gramaticales. La comparativa entre los valores absolutos y

la frecuencia relativa de dichas categorías permite entenderlas como rasgos de estilo autorial, aunque parece que con una señal menos notable que en el componente léxico. Por ejemplo, se dedican varias páginas (102-109) a presentar diferentes tipos de análisis morfológico con cuatro novelas decimonónicas de dos hombres y dos mujeres con la intención de tratar de detectar patrones que permitan la distinción de género. Los resultados son, en este punto, limitados, con la preferencia por coordinación o subordinación como distinción más notable. Sin embargo, como confiesan los mismos autores, un corpus tan reducido no basta para extraer conclusiones extrapolables para un rasgo complejo como es el género del autor. Por otro lado, la coincidencia –más bien léxica, en realidad– de los sustantivos más usados entre hombres y mujeres refleja más bien las tendencias temáticas de autores y autoras del período, más que un estilo genérico propiamente dicho. Sin embargo, los ejemplos sugieren que el análisis morfológico, con un corpus más amplio y un estudio más detallado, puede ofrecer resultados significativos para identificar rasgos estilísticos. Interesantes también pueden ser la detección y el análisis de fenómenos de prefijación y derivación, pues, al cuantificarse y sistematizarse, permiten, una vez más, detectar diferencias notables entre autores.

El sexto capítulo es, junto con el cuarto, el más interesante por la variedad de enfoques y ejemplos que ofrece. Está dedicado al análisis de la sintaxis desde la estilometría, aunque, en el fondo, interviene de nuevo el elemento léxico. Así, si el número de palabras por frase afecta a la legibilidad de un texto y a su menor o

mayor complejidad, los fenómenos de las colocaciones –dos o más palabras seguidas que se repiten– o las coocurrencias –dos palabras que tienden a coincidir en un contexto relativamente pequeño– combinan sintaxis –sintagmas y su organización– con el sentido de las palabras empleadas. Con ejemplos tomados en este caso de programas políticos, se enseña a leer los resultados cuantitativos de la estilometría para detectar tendencias cualitativas. Abordar los marcadores del discurso puede ayudar a “predecir la orientación semántica del texto” (132), lo que tal vez pueda adaptarse al futuro para analizar sistemáticamente el desarrollo lógico-semántico de novelas o ensayos de épocas diferentes. Con todo, como señalan los autores, hay fenómenos interesantes que escapan –al menos por ahora– de la cuantificación de la estilometría, como la ironía, o que apenas dejan huella en la estructura sintáctica de los textos.

El campo del análisis semántico desde la estilometría, al que se dedica el capítulo séptimo, es el que tiene más perspectivas de crecimiento a medio y largo plazo gracias a las posibilidades que aportan técnicas de aprendizaje automático para etiquetar correctamente redes semánticas en grandes cantidades de textos. Las páginas dedicadas a la semántica ilustran algunas posibilidades que ya ofrece la estilometría hoy en día, desde la identificación de conceptos más característicos de un texto para localizar los campos semánticos con los que se relacionan y cómo se diferencian del resto del corpus utilizado, pasando por el contraste cuantitativo del léxico (el ejemplo de dos casos muy específicos, Espronceda y Bécquer, ilustra su utilidad interpretativa),

hasta trazar ciertas tonalidades emocionales dominantes a partir de categorías pre-determinadas. En este último caso, vemos de nuevo que la ausencia de un diccionario de emociones preparado específicamente para el español limita la sutileza necesaria para este tipo de análisis. También presenta limitaciones el modelado de temas o *topic modelling*, que consiste en la identificación automática de los temas de un texto según el léxico dominante, pues los algoritmos existentes todavía no ofrecen resultados muy fiables sin una supervisión humana.

El libro acaba con unas conclusiones generales de las ideas presentadas a lo largo de las páginas precedentes y la bibliografía pertinente, pero también útiles apéndices en forma de un índice onomástico y temático, las fuentes de donde se han tomado los textos analizados (se supone que están disponibles en un repositorio de Github, aunque el enlace proporcionado no parece estar activo) y un glosario de términos vinculados a la estilometría.

Vuelvo a la pregunta inicial con la que comencé esta reseña: ¿quién teme a la estilometría? Quien no sepa qué es y para qué sirve a los investigadores en humanidades. Este libro, único en el ámbito en español, busca quitar este temor explicando y ejemplificando cómo funciona la estilometría aplicada a textos en español. La estilometría está aquí para quedarse, pues supone un haz de análisis cuantitativos que son ya imprescindibles para quienes quieran explorar cuestiones de estilo y lo que conlleva: autoría, género, datación, ideología, etc. Somos los investigadores quienes debemos seguir experimentando con las posibilidades que ofrece la estilometría para sacarle más provecho, como hasta ahora se

ha hecho sobre todo en relación con problemas de autoría.

Como ya he señalado, este libro no es un manual de uso, sino que cumple la función de explicar con numerosos ejemplos este ámbito de las Humanidades Digitales a quien lo desconozca. Con un poco de suerte, animará a más de uno a adentrarse en el uso de unas herramientas que, si antaño parecían muy lejanas a los recursos de análisis filológico, hoy deben formar parte de nuestra paleta metodológica. Una de las limitaciones del libro es que los ejemplos se basan en corpus pequeños de textos, cuando la capacidad de trabajar con grandes cantidades de textos es el punto fuerte de las herramientas digitales. El uso de corpus amplios y ambiciosos (la obra completa de un autor, décadas de obras pertenecientes a un subgénero literario) despliega todo el potencial de la estilometría digital para responder preguntas complejas. El análisis de grandes cantidades de obras ya ha mostrado importantes resultados (cf., por ejemplo, el libro de José Calvo Tello sobre la novela en la Edad de Plata) y será especialmente productivo en futuros estudios de literatura en español. A fin de cuentas, conocer y entender lo que hoy se puede hacer es lo que permite imaginar nuevas preguntas para el mañana.

DOI 10.14672/1.2024.2486