

Plecháč, Petr. *Versification and Authorship Attribution*. Praga, Karolinum/ICL, 97 págs.

Si algo se puede afirmar sobre el libro de Petr Plecháč,¹ es que demuestra las enormes posibilidades que ofrecen las nuevas herramientas tecnológicas para resolver viejos problemas del campo de los estudios literarios. *Versification and Authorship Attribution* describe la evolución de la estilometría, discute los avances más recientes de esta rama y realiza una serie de experimentos a partir de ellos. En la medida en que, en palabras de Plecháč, el objetivo del libro es “examinar la aplicabilidad de las características de la versificación a proyectos de atribución de autoría”(7),² estos experimentos incluyen herramientas de la versología como parte central de su método.

Los estudios estilométricos para la atribución de autoría a través de métodos cuantitativos llevan más de un siglo de desarrollo. Estos se basan en identificar la frecuencia de algunos elementos textuales —como la longitud de las palabras o la cantidad de veces que aparecen algunas de ellas en el texto— para describir el estilo de un autor. Franco Moretti se adhiere a estos métodos y los apoya teóricamente cuando afirma que: “el estilo siempre es una cuestión de frecuencia relativa” (184). En la introducción de *Versification and Authorship Attribution*, Plecháč señala que este campo ha evolucionado en torno a la prosa. Por ello, los métodos más sofisticados para la atribución de autor, basados en *machine learning*, únicamente se han usado para el análisis de datos de naturaleza lexical.

1 Petr Plecháč es un filólogo checo, magíster en filología de la Universidad de Palacký, doctor en teoría de la literatura de la misma universidad y doctor en lingüística matemática de la Universidad Carolina. Trabaja en el Instituto de Literatura Checa, es miembro del Círculo lingüístico de Praga desde 2013 y de la Asociación de estudios eslavos, eurasiáticos y de Europa oriental desde 2014. Ha realizado varias investigaciones a partir de métodos de análisis cuantitativo del verso. El trabajo expuesto en *Versification and Authorship Attribution* es un desarrollo más profundo de algunos problemas tratados en su publicación de 2018: *Versification and authorship attribution. Pilot study on Czech, German, Spanish, and English poetry*.

2 Todas las traducciones del inglés son nuestras.



Buena parte de las características concretas del verso —como el metro o el esquema de la rima— obedece a la tradición y al género, y son deliberadas, por lo que no permite distinguir el estilo de un autor de otro. De ese modo, parece que el estudio de autoría en textos en verso no podría hacerse con los mismos métodos que se usan en la prosa. No obstante, de acuerdo con Plecháč, existen algunas características de la versificación que se escapan del control racional del autor, por lo que son propias de su estilo, como “la preferencia por ciertas disposiciones rítmicas y las frecuencias de los sonidos en la rima” (7). En ese orden de ideas, el libro resalta la importancia de considerar características versológicas para describir el estilo de un autor.

Plecháč realiza una exposición de la evolución de los métodos cuantitativos aplicados a la atribución de autoría a partir de las limitaciones que encuentra en cada uno de ellos. Así, aborda el desarrollo del campo de conocimiento desde dos perspectivas: la de la naturaleza de los datos estudiados y la del manejo estadístico de estos. Su exposición empieza con la descripción del método usado por Thomas Corwin Mendenhall (1901) quien realizó un estudio comparativo del estilo de Shakespeare, Bacon y Marlowe a partir de la frecuencia de la longitud de las palabras más repetidas en algunas de sus obras literarias. De este experimento surge la certeza de que Shakespeare usa más palabras de cuatro caracteres que de cualquier otra longitud, mientras que en Bacon son más frecuentes las palabras de tres caracteres.

El *delta* de Burrow, herramienta propuesta al inicio de los años 2000, continúa esta tendencia en los análisis cuantitativos: propone comparar los estilos de los autores a partir de la frecuencia con la que aparecen las palabras más repetidas en el texto, como artículos y preposiciones.³ Este método abrió paso al análisis multivariado de características estilísticas. En la primera década de los 2000, se desarrollaron algunas herramientas de aprendizaje automático o *machine learning* que facilitaron los análisis estadísticos con

-
- 3 La herramienta se utiliza con el siguiente método. En primer lugar, se escoge la obra de autoría dudosa y un *corpus* de obras con autores definidos, y se determinan las palabras más comunes. Después, se calcula la frecuencia relativa de cada palabra y se representa como un vector. En tercer lugar, se convierte dicha frecuencia relativa en un *z-score*, con la finalidad de que los datos sean comparables entre sí. Por último, se halla el promedio de la diferencia entre los *z-score*. Entre menor sea el valor del *delta* de Burrow, el cual corresponde al promedio de la diferencia entre los valores del *z-score*, habrá mayor similitud entre el texto objetivo y las obras con las que se compara.

múltiples variables. Herramientas como el *support-vector-machine* hacen posible, por ejemplo, determinar qué variables son más significativas para poder distinguir cuantitativamente el estilo de un autor del de otro.

Antes del estudio de Plecháč, ya se habían llevado a cabo investigaciones a partir de elementos de la versología. En 2014, Marina Tarlinskaja publicó el libro *Shakespeare and the Versification of English Drama, 1561–1642*, en que trata con el problema de la autoría de *Enrique VIII*, obra escrita “a cuatro manos” entre Shakespeare y Fletcher. La variable que utiliza para distinguir el estilo de los dos autores es la frecuencia con que ocurren cortes sintácticos fuertes (*strong syntactic breaks*) en las diferentes posiciones métricas. De ese modo, mientras en Shakespeare los cortes sintácticos fuertes ocurren con más frecuencia en la décima sílaba y en la sexta del pentámetro, en Fletcher son más frecuentes en la posición diez y en la siete.

Plecháč considera insuficiente el estudio de Tarlinskaja por tener en cuenta una sola variable y, en su investigación, propone incluir tres características del verso: el ritmo, la rima y la eufonía, además de variables lexicales de *n*-gramas.⁴ Tras el estudio con tres *corpus* de tamaño considerable —uno en castellano, otro en checo y otro en alemán— Plecháč concluye que, si bien los métodos que toman en cuenta variables de la versificación son más precisos que los métodos lexicales, lo ideal para poder llegar a la mayor precisión es utilizar variables de las dos naturalezas.

Una de las mayores contribuciones de su investigación al campo de la estilometría es la forma en la que trabaja con los elementos de la estructura versal. Plecháč afirma que el ritmo, en cuanto realización del metro, no depende de un proceso racional consciente y, por lo tanto, es un rasgo propio del estilo de cada autor. Además, propone estudiarlo desde dos enfoques metodológicos: la frecuencia de los tipos de ritmo en el verso y la frecuencia de los patrones rítmicos en subsecciones del verso. Este último recibe el nombre de *Rhythmic N-grams* y se basa en dividir el verso en porciones de diferentes tamaños, de modo que disminuye la dispersión de los datos en el análisis estadístico. Este método tiene especial aplicabilidad en sistemas de versificación silábicos y silabotónicos, razón por la cual, en la tercera parte

4 La división por *n*-gramas es un método en el que se mide la frecuencia de repetición de subconjuntos rítmicos.

del libro, Plecháč lo utiliza para el análisis de los *corpus* en checo y castellano, mientras que para analizar el *corpus* en alemán, predominantemente tónico, utiliza el método de la frecuencia de los tipos de ritmo en el verso.

Al igual que el ritmo, Plecháč considera la rima como una característica distintiva del estilo de un autor y realiza su análisis a partir de siete factores morfológicos y fonológicos en los pares de palabras rimantes. Por último, la eufonía, que se define como la predilección por la repetición de ciertos sonidos, se aborda en toda la muestra y no únicamente dentro de un solo verso, como es habitual en esta clase de análisis.

Aunque el eje central del estudio sea el tomar en cuenta las variables propias del verso para hacer la atribución autorial, Plecháč resalta que la combinación de métodos lexicales y de versificación arroja resultados más precisos. En la cuarta parte del libro, el autor aplica un método que tiene en cuenta variables de naturaleza mixta en el estudio de *The Two Noble Kinsmen* (obra resultante de la colaboración entre Fletcher y Shakespeare). Los filólogos comenzaron a estudiar el problema de la autoría de sus partes en 1812, cuando únicamente se utilizaban herramientas lexicales. En ese entonces, se tomaba en consideración la repetición de palabras poco usuales y de algunas metáforas. Hacia el año 1993, algunos investigadores utilizaron redes neuronales artificiales para encontrar la frecuencia de repetición de ciertas palabras en obras de los posibles autores. No fue sino hasta 2014, con investigadoras como Marina Tarlinskaja, que las herramientas de la versología empezaron a ser aplicadas a este problema.

La investigación de Plecháč empieza con la separación de obras de los dos autores por escenas, como se había hecho en investigaciones previas. Las muestras fueron analizadas a partir de tres modelos: el que está ligado a un estudio versológico, el lexical y la combinación de ambos, como se muestra en la siguiente tabla:⁵

5 Para modelos basados en 1) los quinientos tipos más comunes de ritmo, 2) las quinientas palabras más comunes y 3) vector de mil dimensiones combinando características 1) y 2). Las cifras muestran la proporción de escenas correctamente clasificadas a lo largo de las trescientas iteraciones. Tabla tomada de Plecháč 73.

Tabla 4.2. Precisión de la identificación de autor

		Rhyt. type-based models	Word-based models	Combination models
Shakespeare	<i>Coriolanus</i>	1	1	1
	<i>Cymbeline</i>	0.997	1	1
	<i>The Winter's Tale</i>	1	1	1
	<i>The Tempest</i>	1	1	1
	<i>Valentinian</i>	0.992	1	1
Fletcher	<i>Monsieur Thomas</i>	1	1	1
	<i>The Woman's Prize</i>	1	1	1
	<i>Bonduca</i>	1	0.93	1

En esta tabla se mide la precisión, en términos de frecuencia relativa, de los tres modelos para atribuir correctamente la autoría de distintas obras de Shakespeare y Fletcher. En este caso, 1 equivale a una precisión total, y todo lo que se encuentra por debajo de él constituye errores en el proceso.⁶ El modelo de Plecháč demuestra que la combinación de métodos suprime los errores que resultan de la aplicación de uno solo uno de ellos. Al ingresar muestras de algunas escenas de cuatro obras de Shakespeare y cuatro obras de Fletcher, el método que utiliza el análisis de los patrones rítmicos tuvo un error de 0,003 en el caso de *Cymbeline* y de 0,008 al atribuir algunos patrones rítmicos presentes en *Valentinian* a Shakespeare. En el caso del análisis de la repetición de palabras, se presentó un error de 0.07 en la atribución del quinto acto de *Bonduca*.

No obstante, al momento de comparar los resultados obtenidos con *The Two Noble Kinsmen*, surgió un nuevo problema: solo la mitad de las escenas eran lo suficientemente largas para ser clasificadas de la misma forma que las muestras anteriores. Esto abrió paso a un nuevo avance en la investigación de Plecháč: la nueva técnica, inspirada en un método utilizado por Maciej Eder en 2016, recibe el nombre de *rolling attribution*. Allí, la obra fue dividida en grupos de cien versos para conformar las muestras. El uso de la nueva técnica fue comprobado varias veces cambiando las agrupaciones de versos, y, aunque arrojó datos más precisos, dio como resultado final la misma atribución de autoría que la división por escenas.

6 Los errores ocurren cuando la aplicación de alguno de los métodos resulta en la atribución de una de las obras al autor equivocado.

El último objeto de estudio del libro de Plecháč, en coautoría con el investigador Artjoms Šeļa, es el caso de los poemas atribuidos a Gavriil Stepanovich Batenkov, poeta y militar ruso nacido en 1793. En 1978 apareció un estudio de su poesía titulado *Poesía del decembrista Gavriil Stepanovich Batenkov* al que se encuentra adjunta una colección de sus poemas inéditos. El autor del libro es Aleksandr Iliushin, versólogo, profesor universitario ruso, y poeta conocido por sus juegos literarios e imitaciones. En 1997 y 1998, Maksim Shapir publica dos investigaciones que ponen en duda la autenticidad de la autoría de la colección editada por Iliushin. El investigador analiza cuidadosamente varios niveles lingüísticos del estilo de los poemas usando métodos cuantitativos, de modo que consigue identificar diferencias significativas entre los poemas publicados en 1978 y el resto de la obra conocida de Batenkov. Sin embargo, Shapir concluye que no existe evidencia suficiente para aseverar que esa colección de poemas es apócrifa.

Plecháč y Šeļa afirman que las limitaciones del estudio de Shapir —que toma en consideración una cantidad muy superior de niveles lingüísticos a la de otras investigaciones— se deben a que las características se comparan una a una. Por ello, se propone hacer el análisis multivariado con SVM (*Support-Vector Machines*). Los resultados del experimento apuntan a que, efectivamente, los poemas de la colección de 1973 y el *corpus* de los demás poemas atribuidos a Batenkov no fueron escritos por la misma persona. De ese modo, Plecháč y Šeļa resuelven con evidencia sólida una de las “viejas” incógnitas de atribución de autor.

La relevancia de esta contribución descansa en el hecho de que “si, como argumenta [Shapir], no podemos rastrear la identidad de un autor con base en varios niveles de características lingüísticas, entonces, el concepto de autor, que toma decisiones lingüísticas únicas y reconocibles, no es más que un constructo académico” (Plecháč 81). En esa medida, contar con evidencia cuantitativa que sustenta de forma sólida el carácter único del estilo de cada autor —más allá del género en que se inscribe— permite confirmar el sentido del ejercicio de estudiar las obras literarias como textos escritos por alguien. Este ejercicio, que se ha hecho por intuición en la práctica académica, ahora tiene validez metodológica. Por otro lado, contar con esta evidencia permite validar prácticas extra académicas como la clasificación, promoción y divulgación de un libro alrededor de la figura de su autor en el campo editorial.

Hacia el final del libro, Plecháč llega a la conclusión parcial de que si algo demuestra la estilometría es que el estilo no es algo exclusivo de los grandes escritores: todos los autores tienen un estilo con características únicas a nivel

lingüístico. La propuesta de Plecháč coincide con la apuesta de Franco Moretti por la lectura distante, pues considera que todos los estilos son únicos. Esto indica que hay un panorama literario más amplio que únicamente las obras canónicas, hecho que amplía el objeto de los estudios literarios y de la historia literaria. Pese a la ausencia de un desarrollo específicamente teórico de esta y otras cuestiones, el libro de Plecháč es muy efectivo en sugerir problemas para futuras reflexiones: ¿existe la posibilidad de predecir los cambios en el estilo de un autor a partir de esos métodos?, ¿cómo se podrían utilizar los métodos estadísticos y de *machine learning* para problemas del estudio del verso distintos a la atribución de autor?, ¿la capacidad de la inteligencia artificial para procesar múltiples variables en el estudio del verso abre la posibilidad de encontrar nuevas variables para el análisis estilométrico?, ¿cómo estas herramientas, que permiten comprender los textos escritos en verso a partir de múltiples variables, y el reflejo tangible de dichas variables en las obras, podrían llevar a plantearse nuevas preguntas sobre los fenómenos literarios?

El libro de Plecháč muestra diferentes experimentos con los métodos estadísticos e informáticos más innovadores para el estudio del verso. *Versification and Authorship Attribution* es de naturaleza predominantemente práctica y, en consecuencia, está estructurado en torno a la resolución de problemas en el campo de los estudios literarios que han sido abordados por diferentes investigadores antes que Plecháč. A través de los métodos cuidadosamente estudiados, reestructurados y aplicados al problema de la atribución de autoría de los poemas de Batenkov, Plecháč consigue la evidencia que le hacía falta al estudio de Maksim Shapir, y con ello demuestra la hipótesis del versólogo ruso. Además, a través del estudio de casos concretos, aporta evidencia sobre la validez de las prácticas académicas vinculadas a la estilometría en general. La lectura de *Versification and Authorship Attribution* resulta especialmente provechosa para acercarse a los estudios cuantitativos del verso, su aplicación a problemas de atribución de autoría y enterarse de los últimos avances tecnológicos y metodológicos en la materia.

Laura Camila Gómez Camelo

Universidad Nacional de Colombia, Bogotá, Colombia

Valeria Muñoz Landinez

Universidad Nacional de Colombia, Bogotá, Colombia

Obras citadas

Moretti, Franco. *Lectura distante*. Buenos Aires, Fondo de Cultura Económica, 2015.