

## La estilometría como técnica de análisis lingüístico-cuantitativo: a propósito del uso terminológico de estilo\*

ISABEL MOYANO MORENO

Universidad de Cádiz (España)

isabel.moyano@uca.es

<https://orcid.org/0000-0003-4284-8897>

Cómo citar:

Moyano Moreno, I. (2023): "La estilometría como técnica de análisis lingüístico-cuantitativo: a propósito del uso terminológico de 'estilo'", *Pragmalingüística*, 31, 339-361. <https://doi.org/10.25267/Pragmalinguistica.2023.i31.15>

### LA ESTILOMETRÍA COMO TÉCNICA DE ANÁLISIS LINGÜÍSTICO-CUANTITATIVO: A PROPÓSITO DEL USO TERMINOLÓGICO DE ESTILO

**RESUMEN:** Los avances en lingüística computacional y de corpus han propiciado un notable impulso de la estilometría como técnica de análisis estadístico aplicada a los corpus textuales. Inserta en el paradigma metodológico de las humanidades digitales, la estilometría es una subdisciplina de la estilística que permite la medición del estilo siguiendo técnicas estadísticas. Dado que la bibliografía sobre este tema es escasa en el ámbito hispano, no encontramos una definición exacta ni consensual de *estilo* como uso terminológico vinculado a este campo. En este trabajo se revisa el término desde distintas perspectivas, considerando su carácter relacional, y se propone un análisis terminográfico con el fin de solventar posibles ambigüedades; de manera paralela, se ha elaborado un corpus especializado en estilometría que facilita el acceso a la información contextual y documental mediante el análisis semi-automatizado.

**PALABRAS CLAVE:** estilometría; atribución de autoría; estilo; terminología; terminografía.

**SUMARIO:** 1. Introducción. 2. La estilometría como técnica de análisis lingüístico-cuantitativo. 3. Breve aproximación teórica a la noción de "estilo"... 4. Metodología. 5. Análisis terminográfico de *estilo* en la estilometría. 6. Conclusiones. 7. Referencias.

**Fecha de recepción:** 13/12/2022

**Fecha de revisión:** 11/01/2023

**Fecha de aceptación:** 27/01/2023

**Fecha de publicación:** 01/12/2023

### STYLOMETRY AS A TECHNIQUE OF LINGUISTIC-QUANTITATIVE ANALYSIS: ON THE TERMINOLOGICAL USE OF STYLE

**ABSTRACT:** Advances in computational and corpus linguistics have given a notable boost to stylometry as a technique of statistical analysis applied to textual corpora. Inserted in the methodological paradigm of digital humanities, stylometry is a sub-discipline of stylistics that allows the measurement of style following statistical techniques. Given that the bibliography on this subject is scarce in the Spanish-speaking world, we have not found an exact or consensual definition of *style* as a terminological usage linked to this field. This paper reviews the term from different perspectives, considering its relational character, and proposes a terminographic analysis in order to solve possible ambiguities; in parallel, a corpus specialised in stylometry has been elaborated to facilitate access to contextual and documentary information by means of semi-automated analysis.

**KEYWORDS:** stylometry; authorship attribution; style; terminology; terminography.

**SUMMARY:** 1. Introduction. 2. Stylometry as a linguistic-quantitative analysis technique. 3. Brief theoretical approach to the notion of "style"... 4. Methodology. 5. Terminographic analysis of *style* in stylometry. 6. Conclusions. 7. References.

### LA STYLOMÉTRIE COMME TECHNIQUE D'ANALYSE LINGUISTIQUE-QUANTITATIVE: À PROPOS DE L'USAGE TERMINOLOGIQUE DE STYLE

**RÉSUMÉ:** Les progrès de la linguistique computationnelle et de la linguistique de corpus ont donné un élan notable à la stylométrie en tant que technique d'analyse statistique appliquée aux corpus textuels. Inscrite dans le paradigme méthodologique des humanités numériques, la stylométrie est une sous-discipline de la stylistique qui permet de mesurer le style à l'aide de techniques statistiques. Étant donné que la bibliographie sur ce sujet est rare dans le monde hispanophone, nous n'avons pas trouvé de définition exacte ou consensuelle de *style* en tant qu'usage terminologique lié à ce domaine. Cet article examine le terme sous différentes perspectives, en tenant compte de son caractère relationnel, et propose une analyse terminographique afin de résoudre les éventuelles ambiguïtés; parallèlement, un corpus spécialisé en stylométrie a été élaboré pour faciliter l'accès à l'information contextuelle et documentaire au moyen d'une analyse semi-automatisée.

**MOTS-CLÉS:** stylométrie; identification de l'auteur; style; terminologie; terminographie.

**SOMMAIRE:** 1. Introduction. 2. La Stylo-métrie comme technique d'analyse linguistique-quantitative. 3. Brève approche théorique de la notion de "style"... 4. Méthodologie. 5. Analyse terminographique de *style* en stylométrie. 6. Conclusions. 7. Références.

\* Esta investigación se inscribe en el proyecto "Aplicaciones de la lingüística digital al ámbito de la terminología: la creación de un léxico relacional bilingüe de los usos terminológicos de la semántica léxica (TerLexWeb)" (PID2022-139201OB-I00), financiado por la Agencia Estatal de Investigación del Ministerio de Ciencia e Innovación.

## 1. Introducción

Las humanidades digitales son un conjunto de metodologías que establecen un puente entre lo humanístico y lo digital, permitiendo que aquellos estudios que hasta hace relativamente poco estaban relegados a un tratamiento cualitativo puedan beneficiarse de técnicas de análisis cuantitativas (Crespo Miguel, 2021). Su carácter inter/transdisciplinar atrae el interés de distintos campos de conocimiento que convergen en la estilometría, acogiendo fundamentos teóricos naturales de la estilística y técnicas propias de la estadística, la lingüística de corpus y la lingüística computacional. Como una de las ramas más asentadas y con mayor proliferación en el ámbito de las humanidades digitales, la estilometría es una subdisciplina de la estilística lingüística (Wales, 2014b: 402) que propone técnicas de análisis enfocadas en el estudio de la relación que existe entre el estilo de escritura de los textos y los metadatos contenidos en dichos textos (i.e. autoría, género textual, etc.) (Boto Bravo, 2017).

La estilometría es un nuevo ámbito especializado que surge gracias a la sinergia entre lingüística y tecnología (Crespo Miguel, 2021) y la implementación de métodos informáticos en humanidades; sin embargo, carece de una variedad de obras especializadas que versen sobre ella, especialmente en el ámbito académico hispanoparlante. Un problema general que envuelve a los ámbitos especializados y al que se enfrentan los expertos de distintos campos del saber es la unificación y organización de sus conceptos, sobre todo en aquellas áreas de especialidad aún en desarrollo. Desde finales del siglo pasado, la ciencia terminológica ha presenciado una transición de una terminología prescriptiva hacia una terminología descriptiva. En otras palabras, la terminología moderna ya no busca el establecimiento de correspondencias biunívocas entre concepto y denominación lingüística, sino recopilar y describir las unidades terminológicas de las lenguas de especialidad, teniendo en cuenta cuestiones de variación conceptual y formal; esto se traduce en el desarrollo de distintos recursos terminológicos de carácter descriptivo que consignen y analicen las unidades terminológicas en glosarios o léxicos especializados (Ramírez Salado, 2019).

En el caso de la lingüística, su asentamiento como ciencia, la proliferación en líneas de investigación y publicaciones y, en general, la cantidad masiva de fuentes y datos contenidos en corpus han desembocado en una gran dispersión y variación terminológica. En el Instituto Universitario de Lingüística Aplicada (ILA) de la Universidad de Cádiz se vienen desarrollando en los últimos años proyectos centrados en el estudio descriptivo de la terminología lingüística y la extracción, tanto manual como semiautomática, de usos terminológicos contenidos en corpus de carácter metalingüístico. Desde el proyecto de excelencia del

MINECO “Comunicación especializada y terminografía: usos terminológicos relacionados con los contenidos y perspectivas actuales de la semántica léxica (TERLEX)<sup>1</sup>” (Casas Gómez, 2015) se ha resaltado la necesidad por la digitalización de fuentes lexicográficas especializadas en múltiples lenguas y nuevas publicaciones en el ámbito de la lingüística para la construcción de un corpus metalingüístico y multilingüe que sirva como corpus de referencia y como fuente para la extracción de usos terminológicos vinculados a los contenidos y perspectivas actuales de la lingüística y, en concreto, de la semántica léxica.

La definición de *estilo* es una cuestión central para los especialistas de la estilometría. Pese a que este concepto abarca varias áreas temáticas y ha sido trabajado desde la literatura o desde la estilística, no existe consenso ni una opinión totalmente unívoca al respecto vinculada a una óptica estilométrica. Por ello, este trabajo propone una revisión teórica en torno a la noción de “estilo” desde distintas perspectivas, así como la elaboración de una ficha terminológica (Casas Gómez, 2020, 2022) sobre el uso de este término en el contexto técnico de la estilometría con el fin de resolver posibles cuestiones de ambigüedad y también contribuir al estatus científico de la estilometría. Para su desarrollo se plantean los siguientes objetivos:

1. presentar la estilometría como subdisciplina de la estilística y técnica de análisis lingüístico-cuantitativo inserta en el paradigma metodológico de las humanidades digitales;
  2. ofrecer, a partir de la revisión bibliográfica, una definición del término *estilo* enmarcada en el campo de la estilometría; y
  3. proponer una ficha terminológica de *estilo* como uso técnico vinculado a la estilometría, partiendo de la recopilación de un corpus especializado cuyas fuentes servirán como referencia y soporte para el trabajo terminográfico (cf. 4.1.).
- 2. La estilometría como técnica de análisis lingüístico-cuantitativo**

La naturaleza ecléctica e interdisciplinar de la estilística la convierte en una disciplina abierta a un amplio abanico de metodologías (Wales, 2014b; Orgoñová, 2019). En su dimensión lingüística se encuentra la posibilidad de una estilística a gran escala basada en corpus (Wales, 2014b: 44) enfocada a encontrar patrones lingüísticos que no serían discernibles por el ojo humano. Como su propia etimología anticipa, la estilometría consiste en la medición del estilo. Estudios pioneros que datan de finales de siglo XIX sitúan una primera consideración en torno a la posibilidad de medir el estilo (Holmes, 1998),

---

<sup>1</sup> <https://ila.uca.es/proyectos/>

aunque podemos establecer unos antecedentes historiográficos desde la lexicometría propia de la lingüística francesa (Dugast, 1980; Ménard, 1983; Allen, 1984), técnica que ya anticipa la posibilidad de medir patrones lingüísticos y el tratamiento cuantitativo del léxico. Pese a que lexicometría y estilometría comparten antecedentes similares y métodos convergentes, divergen, sin embargo, en su objeto de estudio: la lexicometría se centra en el análisis de frecuencias de las unidades léxicas (i.e. las palabras), mientras que la estilometría se centra en la medición del estilo, acogiendo el análisis de frecuencias léxicas en algunos casos, pero considerando otros muchos rasgos de estilo. Una primera aparición del término *estilometría* se da a finales del siglo XIX (Lutosławski, 1898; Hernández Lorenzo, 2022); sin embargo, estos primeros estudios no resultan más que un esbozo de lo que supone la estilometría actualmente, gracias a la aparición y al avance de soportes computacionales para el procesamiento y el análisis masivo de datos mediante una lente macroscópica o una lectura distante<sup>2</sup>. Desde entonces hasta la actualidad, la estilometría ha ido beneficiándose paulatinamente de las posibilidades ofrecidas por los métodos estadísticos e informáticos para resolver cuestiones de atribución de autoría o incluso realizar tareas propias del análisis literario.

Actualmente, la estilometría se considera una técnica de análisis lingüístico-estadístico (Lozano Zahonero, 2020), resultado de la sinergia entre la estilística y la lingüística informática (Eder y Kestemont, 2016; Boto Bravo, 2017), que se realiza sobre corpus textuales. Este tipo de técnica no tiene como objetivo sustituir los métodos asentados en la estilística tradicional, sino más bien aportar nuevos enfoques ayudándose de métodos informáticos. La estilometría basada en corpus permite la replicabilidad y el trabajo experimental o cuantitativo, evitando afirmaciones imprecisas para llegar, en su lugar, a evidencias empíricas:

What, then, does stylometry seek to do? It certainly does not seek to overturn traditional scholarship by literary experts and historians, rather it seeks to complement their work by providing an alternative means of investigating (...). (Holmes, 1998: 111).

El propio término *estilometría* ha adquirido diferentes denominaciones lingüísticas a lo largo de los años y según diferentes autores, especialmente en sus inicios, cuando aún quedaba por determinar la

<sup>2</sup> Dos términos clave en el estudio de las humanidades digitales son el *close reading* y el *distant reading*, este último introducido por Moretti (2000, 2013). El primero se corresponde con la lectura y el análisis de textos tradicional que ofrece la crítica literaria, mientras que la segunda, fundamentada en técnicas cuantitativas, permite una lectura a gran escala, es decir, el acceso y el tratamiento inmediato de colecciones de textos. Esta distinción entre lectura microscópica y lectura macroscópica ha sido desarrollada y complementada por otros autores (Jockers, 2013).

interfaz entre estilística y estadística: *estilística estadística* (Bailey y Doležel, 1968), *estilística cuantitativa* (Arnaud, 1976), *estiloestadística* (Weiger, 1980), *estilística computacional* (Kock *et al.*, 1983) o *estudios estadístico-estilísticos* (Allen, 1984). Esta variedad de denominaciones revela que, efectivamente, la estilometría se nutre en gran parte de la estilística lingüística, disciplina de larga tradición en la lingüística francesa y general (Bally, 1909 [1970]; Riffaterre, 1971), con el matiz de que la estilística tradicional se basa en un análisis cualitativo e, inevitablemente, subjetivo, mientras que la estilometría es esencialmente cuantitativa y empírica. Al realizar el análisis de corpus a través de métodos estadísticos, la estilometría permite que el investigador consiga apoyar sus hipótesis iniciales en datos puramente empíricos y objetivos. El texto se descompone entonces en un conjunto de datos extraíbles y analizables y facilita estos resultados al estilometrista, que encuentra en la repetición de determinados elementos una serie de patrones que pueden ser característicos o particulares de un autor (Fernández Trillo, 2015: 4). De esta manera, el fin del análisis estilométrico es la detección de patrones lingüísticos que no podrían ser percibidos mediante métodos tradicionales o mediante una lectura cercana (Lozano Zahonero, 2020).

Si bien suele categorizarse como una disciplina de naturaleza forense, dada la relevancia y los avances en aplicaciones como la atribución de autoría o la detección de plagio (Queralt, 2020), el análisis estilométrico no es exclusivo de esta materia, ya que sus técnicas son “eminente mente estadísticas” (Ramírez Salado, 2019: 204) y aplicables a distintas ramas del conocimiento. Como parte de la estilística lingüística, el estudio del estilo, independientemente de si es cualitativo o cuantitativo, es esencialmente lingüístico. Hay autores que apuntan que la estilometría no deja de ser la estadística “al servicio de la lingüística” (Fernández Trillo, 2015).

Stylistics has been defined as a sub-discipline of linguistics that is concerned with the systematic analysis of style in language and how this can vary according to such factors as, for example, genre, context, historical period and author. (Jeffries y McIntyre, 2010: 1).

El análisis de estilo es esencialmente contrastivo, es decir, que el estilo de un autor viene caracterizado por los rasgos o marcadores de estilo que lo diferencian de otros. Existen técnicas propias de la estadística para el trabajo contrastivo de estilo (entre ellas *chi cuadrado* o *log likelihood*) que consiguen extraer los rasgos de estilo de un autor que resultan estadísticamente prominentes, seleccionando únicamente aquellos rasgos de estilo significativos que suponen una diferencia estadística con respecto a otros autores. En los estudios de estilometría actuales esto implica el trabajo basado en corpus y

las técnicas propiamente desarrolladas desde la lingüística de corpus (Fuster Márquez, 2022), que dotan al investigador de evidencias empíricas y datos objetivos acerca de usos reales producidos de manera natural, con el fin de identificar patrones lingüísticos en un contexto determinado (Carrió Pastor, 2020). Ahora bien, la supuesta objetividad sobre la que se asienta la estilometría no es del todo cierta, ya que la investigación puede verse sesgada por la elección de unos rasgos de estilo u otros<sup>3</sup>, por lo que la labor humana prevalece tanto en una fase previa (i.e. la selección de los rasgos que se van a analizar) como en una última fase de interpretación y evaluación de los resultados. El análisis estilométrico puede realizarse bien desde enfoques digitales, empleando plataformas en línea para la visualización macroscópica de fenómenos lingüísticos (Crespo Miguel, 2020: 152) como *Voyant Tools*<sup>4</sup> o *Sketch Engine*<sup>5</sup>, que no requieren conocimientos previos ni procedimientos en programación, sino que permiten la carga de un corpus para la elaboración de nubes de palabras, listados de frecuencias y gráficos que pueden arrojar luz sobre cuestiones de estilo; o bien desde enfoques computacionales propios del procesamiento del lenguaje natural, empleando lenguajes y paquetes de programación centrados especialmente en la praxis estilométrica, como es el caso del paquete *Stylo*<sup>6</sup> (Eder y Kestemont, 2016) en lenguaje de programación R, especialmente diseñado para el análisis estilométrico de corpus mediante el análisis de distancias.

En cuanto a las tareas principales de la estilometría (Neal *et al.*, 2017) encontramos el análisis de autoría forense, en el que se insertan la atribución y verificación de autoría, el análisis de estilo, la elaboración de perfiles lingüísticos y la detección de plagio, tareas asociadas al lenguaje escrito y el reconocimiento de locutores, vinculado al lenguaje oral. No obstante, según apunta la bibliografía científica, aquellas aplicaciones que se integran en el lenguaje escrito están empezando a diversificarse, dejando a un lado su componente forense y atrayendo a especialistas de, por ejemplo, la crítica literaria, que comienzan a explorar lo que este tipo de análisis y las herramientas diseñadas para llevarlo a cabo pueden ofrecer a sus estudios, como es el caso de obras pioneras en el ámbito académico anglosajón (Horton, 1987) y más recientemente en el hispano (Lozano Zahonero, 2020; Vega García-Luengos, 2021).

<sup>3</sup> En el ámbito de la estilometría se han llevado a cabo estudios de carácter evaluativo en torno a los marcadores de estilo más significativos y relevantes (Holmes, 1998), así como el desarrollo de clasificaciones taxonómicas de los mismos (Grieve, 2007; Stamatatos, 2009; Neal *et al.*, 2017).

<sup>4</sup> <https://voyant-tools.org/>

<sup>5</sup> <https://www.sketchengine.eu/>

<sup>6</sup> El paquete *Stylo* se encuentra disponible para descarga en el sitio web oficial [www.r-project.org](http://www.r-project.org).

### 3. Breve aproximación teórica a la noción de “estilo” desde la estilística lingüística hasta la estilometría

El análisis estilométrico parte de la premisa de que todas las personas hacen uso del lenguaje de un modo distinto e idiosincrásico (Coulthard *et al.*, 2016), sosteniendo que todo autor, de manera inconsciente, tiene su propia huella lingüística que lo distingue de otros autores (Daelemans, 2013). La presuposición de que todo individuo tiene su propio estilo desemboca en planteamientos y reflexiones teóricas en torno al estilo. No obstante, la noción de estilo es de definición heterogénea y relacional, ya que un mismo concepto puede ser definido desde distintas perspectivas y terminar adoptando una terminología u otra según la aplicación a la que esté enfocado. Es más, *estilo* es un término cuya definición no solo es complicada (Akhmanova, 1970) sino que además ha sido ampliamente estudiada y explotada, sometida a todo tipo de excesos, tanto en estudios lingüísticos, como literarios (Olsson, 2008). Hay varias áreas interrelacionadas e implicadas en la definición de *estilo* (Wales, 2014a: 397). En este trabajo, las que nos ocupan son aquellas que se encuentran subordinadas a la lingüística.

A grandes rasgos, una aproximación general en torno al estilo la encontramos desde la estilística lingüística, disciplina a caballo entre la lingüística y la psicología (Bally 1909 [1970]: 13) dedicada al estudio de los valores afectivos del lenguaje y la manifestación del pensamiento a través del mismo. Bally establece una división tripartita del estilo que parte de esta distinción *lengua/habla*: un estilo del lenguaje general, otro que concierne a las lenguas particulares y, por último, uno que estudia el habla individual, que subdivide en dos tipos: el estilo individual y el estilo de un autor u orador, concebido desde una óptica estética o literaria. Es el primer tipo, el estilo individual, el que se corresponde con las primeras bases estructuralistas y las nociones de estilo más cercanas a un idiolecto, puesto que este viene determinado por su diferenciación frente a un grupo más grande. Cada individuo, a su manera, realiza desviaciones con respecto a una gramática, utilizando palabras que los miembros de este otro grupo rara vez usan (Bally, 1909 [1970]: 18). En otras palabras, el estilo según este autor se manifiesta en el uso de determinadas palabras que otros no usan, aunque este también apunta a las limitaciones o dificultades metodológicas del estudio del estilo, advirtiendo que estas diferencias léxicas son, en general, poco aparentes (Bally, 1909 [1970]: 18).

La estilística de Bally se desarrolla en la línea de la clásica dicotomía saussureana de *lengua/habla* y posterior escuela estructuralista praguense, ubicándose el estilo en el ámbito del habla como un uso individual de la lengua más o menos accidental (de Saussure, 1916 [1922]). Pese a que no hay una mención explícita en cuanto al término

en el *Cours de Linguistique Générale* (1916 [1922]), entendemos una referencia al estilo en la materialización del desvío con respecto a la norma que realiza el hablante y la selección propia, bien sea meditada o accidental, de una serie de marcadores específicos sobre un amplio repertorio lingüístico. Esta percepción del estilo se inscribe en el ulterior desarrollo de la interfaz entre semántica y estilística en la semántica “tradicional” de la primera mitad del siglo XX (Casas Gómez, 2009: 140-142; 2021: 77), con autores que tratan el estilo y su interrelación con el significado (Ullmann, 1964, 1973; Baldinger, 1970). Por otra parte, esta concepción de los rasgos de estilo como difícilmente discernibles será, más adelante, uno de los fundamentos básicos de la estilometría.

A finales del siglo pasado, la llamada linguoestilística (Akhmanova, 1970) parte primero de una definición de estilo a grandes rasgos. Este concepto se fundamenta sobre la presuposición de que objetos esencialmente idénticos, que comparten una serie de características primarias, difieren en aspectos secundarios que no forman parte de las propiedades básicas de estos objetos. El concepto de estilo lingüístico (Akhmanova, 1970), no obstante, reside en la manifestación del lenguaje en el habla, coincidiendo con los postulados referentes a que el estilo es la suma recurrente de una serie de patrones lingüísticos.

En la actualidad, desde la lingüística forense, una de las ramas de la lingüística que más ha contribuido a cuestiones referentes al estilo, encontramos la necesidad por definir este término enfocado a uno de los mayores campos de actuación de esta rama de la lingüística aplicada: la atribución de autoría. Esta puede realizarse bien mediante técnicas propias de la acústica forense, cuando se trata de grabaciones de autoría dudosa, o bien mediante métodos propios de la estilística lingüística, cuando se trata del análisis de corpus textuales. Una primera aparición del mismo término *lingüística forense* aparece vinculado a la estilística (Crystal, 1997; Ramírez Salado, 2019); por ello, podemos contemplar el estilo desde la lente de una estilística forense, como recoge Ramírez Salado (2019: 283) con una definición del estilo como el peritaje lingüístico enfocado a la atribución de autoría, considerando una serie de rasgos o indicadores aislados para medir las diferencias entre un conjunto de autor/es y otro/s. Según apunta la bibliografía científica especializada en lingüística forense y en estudios de estilística forense contrastiva, se suele optar por el término *idiolecto* o *estilo idiolectal* como uso técnico asociado a esta materia; en ocasiones, esta decisión terminológica deriva en la preferencia del uso de *idiolectometría* (Ramírez Salado, 2019: 326-328) sobre *estilometría*. En términos generales, la bibliografía apunta a que la distinción de rasgos de estilo suele ser más sencilla cuando se trata de un texto oral, ya que en el nivel fonológico se pueden apreciar grandes huellas distintivas (Fernández Trillo, 2015: 4). No encontramos, sin embargo, estos

rasgos fonológicos distintivos en un texto escrito, pero en cierto modo se da una combinación de selecciones conscientes e inconscientes de los recursos lingüísticos que se van a emplear a la hora de modelar el mensaje. Al final, el estilo a la hora de escribir un texto consiste en una combinación de decisiones aplicadas a distintos niveles lingüísticos (Daelemans, 2013), como ya exponía de Saussure (1916 [1922]). Estas decisiones, conscientes o no, desembocan en la formulación de la hipótesis del estilema humano, cuya base es que el estilo de un autor se trata de algo intrínseco e individual, casi como si se tratase de una huella dactilar (Daelemans, 2013: 451). Esta hipótesis puede ser en ocasiones descartable o al menos objeto de duda, ya que se plantea una cuestión clave (Daelemans, 2013) pendiente de resolución en el campo de la estilometría en lo referente a la definición de estilo: ¿es el estilo algo invariable o es propenso a una evolución con el paso del tiempo? Lo cierto es que, pese a que la escritura de un individuo está sujeta a cierto cambio y evolución a lo largo del tiempo, los estudios en estilística forense insisten en que hay algo que permanece estático y constante en la producción escrita (Wales, 2014a: 402; Fernández Trillo, 2015: 49).

Por último, desde el punto de vista de la estilística de corpus y la estilística computacional, enfoque más aproximado a la estilometría tal y como la conocemos, se asume que un autor selecciona y utiliza un número finito de palabras (de nuevo, observamos la consideración del léxico como rasgo de estilo) en un contexto determinado. Estos rasgos son los que se analizan y se cuantifican como base para el estudio contrastivo entre conjuntos de corpus textuales. En resumen, el estilo concebido por el tratamiento semi-automatizado del lenguaje viene determinado por la extracción de todas estas decisiones que toma un autor sobre un repertorio lingüístico convertidas en frecuencias (Short, 2018: 330). Habiendo revisado las consideraciones anteriores y los aportes desde otras disciplinas, establecemos una serie de características comunes para construir una definición de *estilo* aplicada a la estilometría:

1. El estilo se enmarca en el habla, considerando la dicotomía de *langue/parole* en el sentido de que es una manifestación individual (el habla) de unos rasgos de estilo seleccionados a partir de un repertorio lingüístico (la lengua).
2. La medición del estilo lingüístico es fundamentalmente contrastiva, pues, para analizar el estilo de un individuo, la producción de este ha de compararse con la de otros autores, a falta de una norma establecida (Wales, 2014a: 402), para observar cuáles son los rasgos que lo hacen destacar sobre otros y, por consiguiente, resultan característicos o particulares de un autor.

3. Los rasgos o marcadores de estilo son aislados y prominentes; el estilo de un individuo se aprecia a partir del análisis de frecuencias de determinados patrones lingüísticos que resultan estables y, pese a que el estilo está sujeto a evolución, hay rasgos de estilo que se mantienen estáticos a lo largo del tiempo.
4. Para llevar a cabo un análisis estadístico de frecuencias, el estilo debe ser cuantificable; por ello, es necesario determinar qué rasgos se van a medir (i.e. la frecuencia de una palabra determinada, la longitud media de las oraciones, ciertas coocurrencias léxicas, etc.).

#### 4. Metodología

En este trabajo se propone un análisis terminográfico a partir de la elaboración de una ficha terminológica (cf. 4.2.) de *estilo* como uso técnico propio de la estilometría. Para desarrollar esta labor terminográfica, planteamos una metodología basada en corpus que se desglosa en las siguientes fases:

1. diseño y compilación de un corpus especializado sobre estilometría;
2. procesamiento del corpus, siguiendo prácticas propias de la lingüística de corpus; y
3. elaboración de la ficha terminológica de *estilo* como uso técnico perteneciente al ámbito de la estilometría.

##### 4.1. CORPUS Y PROCESAMIENTO

El estudio de la terminología basado en corpus es una decisión de corte metodológico que convierte al corpus en una herramienta conveniente para facilitar el desarrollo de la praxis terminográfica, especialmente gracias a las siguientes posibilidades:

1. El vaciado de unidades terminológicas *in vivo* (Santamaría, 2009), es decir, el análisis de los términos en producciones reales.
2. El acceso a la información contextual, la información lingüística y la información sobre fuentes (Díaz Negrillo y Fernández Domínguez, 2010: 38) contenidas en corpus textuales especializados, y
3. Una serie de facilidades metodológicas (i.e., la posibilidad de crear listas de palabras, frecuencias, análisis de coocurrencias léxicas, búsqueda de términos relacionados, etc.) mediante el soporte de medios digitales o computacionales.

Para la labor concreta de elaboración de la ficha terminológica será especialmente útil la segunda cuestión, puesto que el análisis semi-automático de corpus servirá como motor de búsqueda para extraer la información contextual y los contextos definitorios y metalíngüísticos del término propuesto para análisis, así como las fuentes de aparición de cada ocurrencia del término en el corpus, ya que uno de los objetivos principales del trabajo terminográfico es la identificación del conocimiento que envuelve a los términos que aparecen en textos especializados (Alarcón *et al.*, 2007: 247).

Un corpus especializado se define como un tipo de corpus diseñado con el fin de ser representativo del uso lingüístico de un grupo específico de hablantes. La delimitación del criterio de representatividad que envuelve a los lenguajes de especialidad (Pérez Hernández, 2002), sumada al hecho de que la estilometría es un campo de actuación novedoso e interdisciplinar, supone una cuestión problemática de partida en lo referente a una labor previa de recopilación documental. En este trabajo se propone la revisión de estilo vinculada a la estilometría en el ámbito hispánico. Para ello, se ha procedido a la extracción de fuentes especializadas digitales en acceso abierto en español en el motor de búsqueda científico-académico *Google Scholar*, publicadas a partir del año 2010, dado el reciente nacimiento de esta área, y que contengan en el título las palabras *estilometría*, *estilométrico* o *estilométrica*, con el fin de compilar un corpus que suponga una muestra representativa sobre la estilometría en el ámbito académico hispánico. Siguiendo estos criterios, el resultado final es un corpus textual monolingüe de carácter metalíngüístico, confeccionado a partir de 30 fuentes especializadas en estilometría (cf. Apéndice I).

#### 4.2. MODELO DE FICHA TERMINOLÓGICA: *DOCUTERM*

La terminografía, entendida como la dimensión aplicada de la terminología, es la materia encargada de los problemas metodológicos y prácticos con respecto a los lenguajes de especialidad. Esta se ha visto favorecida, especialmente desde finales del siglo pasado, por el soporte de disciplinas transversales como la informática o la documentación (Casas Gómez, 2006, 2022) en lo referente al tratamiento de la terminología y las propuestas de modelos representativos de documentación terminográfica. En concreto, precisamente gracias al diálogo entre la terminología y la documentación, surge la ficha terminológica como uno de los mayores soportes técnicos para la recopilación de unidades terminológicas pertenecientes a un dominio o área de especialidad.

La ficha terminológica *DOCUTERM*, recientemente patentada, (Casas Gómez, 2020, 2022) supone una gran innovación con respecto a

otros modelos de registro terminológico anteriores<sup>7</sup> (Auger y Rousseau, 1977; Cabré, 1993; Dubuc, 1999) y es que, siendo más abarcadora, incluye un mayor número de campos en una ficha técnica de hasta 25 registros<sup>8</sup> y centra el foco de atención en el uso técnico en lugar de en el término, planteando que se recoja una ficha por cada acepción o uso terminológico<sup>9</sup>; de esta manera, los usos terminológicos de un mismo término quedarán consignados bajo un subíndice explicativo (cf. Registro 2), identificando así las expresiones ambiguas que pueden surgir de determinados tecnicismos metalingüísticos. El objetivo final de la elaboración de fichas terminológicas será el de consignar un léxico de especialidad o un glosario de usos técnicos vinculados a los contenidos y perspectivas actuales de un área de especialidad, así como bases de datos y bancos terminológicos electrónicos. En concreto, otro dato que hace que este modelo de ficha resulte tan meritorio es el hecho de que está diseñado para su aplicación a la lingüística, es decir, que está específicamente enfocada a la terminología metalingüística, sirviendo como una herramienta conveniente para cubrir los objetivos y necesidades del proyecto y crear un repertorio terminográfico de carácter metalingüístico.

Además de esta novedad, se incluyen otras como el nivel de metalengua (cf. Registro 12), que determina si el término de análisis se adscribe al metalenguaje general de la lingüística o si, por el contrario, pertenece a alguna escuela o incluso a algún autor en concreto. Casas Gómez (2020) propone, además, la inclusión de ejemplos en diversas lenguas que contribuyan a clarificar los conceptos expuestos (cf. Registro 15) -apartado que no debe confundirse con la traducción del término a otras lenguas (cf. Registro 20)- y la adición de contexto(s) representativo(s) del uso terminológico, que servirán como contextos metalingüísticos de carácter definitorio. Otra de las innovaciones más relevantes añadidas en el modelo de ficha *DOCUTERM* es la consideración de las relaciones conceptuales (de carácter sinonímico, par sinonímico, hiperónimico, hipónimico, merónimico y antónimico) entre usos terminológicos (cf. Registros 16-19). La ficha terminológica y sus correspondientes registros se disponen de manera simplificada en la

<sup>7</sup> Este modelo de ficha terminológica surge como respuesta ante la falta de precisión de otros modelos de ficha anteriores con respecto a los metalenguajes científicos tras una exhaustiva revisión crítica de los mismos (cf. Casas Gómez, 2006).

<sup>8</sup> Los registros de la ficha pueden alterarse según las necesidades del terminólogo. Un ejemplo ilustrativo de esta posibilidad es el trabajo de Ramírez Salado (2019) sobre los usos terminológicos vinculados a los contenidos y perspectivas actuales de la lingüística forense, en el que realiza una serie de modificaciones sobre la ficha original *DOCUTERM* añadiendo un apartado para siglas y una columna extra en la ficha para el relleno de entradas también en inglés con el fin de elaborar un léxico bilingüe español-inglés (Ramírez Salado, 2019: 157-158).

<sup>9</sup> Un ejemplo de la complejidad lingüística que puede suponer un término y su manera de solventar este tipo de problemas mediante el reconocimiento de los usos terminológicos del mismo es el dispuesto por el autor sobre el término *onomasiología* (Casas Gómez, 2022: 17-25).

siguiente tabla, aunque la explicación de cada uno de los registros puede consultarse en la reciente propuesta de modelo de documentación de usos terminológicos metalingüísticos (Casas Gómez, 2022).

1. Término de entrada	2. Subíndice	3. Categoría gramatical
4. Variación denominativa	5. Fuentes del término	6. Estatus
7. Áreas temáticas	8. Etimología	9. Mecanismo de formación
10. Definición	11. Fuente de la definición	12. Nivel de metalengua
13. Contextos representativos	14. Fuentes de cada contexto	15. Ejemplos en otras lenguas
16. Remisiones a términos sinónimos y parasinónimos	17. Remisiones a términos hiperónimos	18. Remisiones a términos hipónimos
19. Remisiones a términos antónimos	20. Coincidencias terminológicas	21. Equivalencias en otras lenguas
22. Fuentes de las equivalencias	23. Fuentes documentales	24. Bibliografía recomendada
25. Observaciones y notas		

Tabla 1. Modelo de ficha terminológica *DOCUTERM*  
y sus 25 registros correspondientes (Casas Gómez, 2020)

## 5. Análisis terminográfico de *estilo* en la estilometría

La ficha terminológica de *estilo* como uso técnico vinculado a la estilometría ha sido realizada alternando métodos manuales y métodos semi-automáticos (cf. Registro 25) basados en corpus. Partiendo del modelo de ficha propuesto (Casas Gómez, 2020, 2022), se ha tomado la decisión metodológica de modificar algunos de sus registros, prescindiendo aquellos registros que atienden a las relaciones conceptuales (cf. Registros 16-19). De manera paralela al análisis semi-automático de contextos definitorios, se ha llevado a cabo una búsqueda manual de equivalencias terminológicas en otras lenguas (cf. Registro 21), ya que trabajamos con un corpus monolingüe y no podemos realizar esta búsqueda siguiendo métodos computacionales.

En primer lugar, se identifica *estilo* como término y *estilo<sub>1</sub>* como único uso terminológico en el campo de la estilometría, ya que no se encuentra el registro de este uso en la base de datos TERLEX, aunque sí se localizan usos de este término asociados a otras disciplinas. Es un término que pertenece a la categoría gramatical de sustantivo, documentado lexicográficamente de manera general adscrito a áreas temáticas como la lingüística y varias subdisciplinas que derivan de ella (i.e. la estilística, la semántica, la sociolingüística, la variación, la

lingüística forense, etc.), pero no específicamente como uso técnico vinculado a la estilometría. La definición propuesta (cf. Registro 9) parte de la revisión bibliográfica de diversas fuentes especializadas en materia de estilometría, contenidas en el corpus de referencia compilado específicamente para la elaboración de esta ficha. Se corresponde con un nivel de metalengua general, es decir, que su uso no pertenece a ninguna escuela concreta o a un uso idiolectal de algún autor en concreto (cf. Registro 11). Por último, añadimos contextos definitorios y metalingüísticos representativos extraídos del corpus de referencia, así como las equivalencias terminológicas en lengua inglesa, francesa y alemana y sus respectivas fuentes. Se adjunta a continuación la ficha con todos sus registros completos:

1. Término de entrada: <i>estilo</i>	2. Subíndice: <i>estilo</i> <sub>1</sub>	3. Categoría gramatical: sustantivo
4. Fuentes del término: corpus de textos especializados y fuentes lexicográficas digitalizadas	5. Estatus: estandarizado, documentado lexicográficamente (cf. Registro 18, Nota 2)	6. Áreas temáticas: estilística, estilística lingüística, estilística computacional, estilística de corpus, estilometría
7. Etimología: del latín: <i>stilus</i> , -i (punzón)	8. Mecanismo de formación: cultismo	
9. Definición:  Definición propia basada en la revisión bibliográfica: el estilo consiste en una serie de decisiones inconscientes que toma un determinado individuo sobre un repertorio lingüístico, sobre el que tiene libertad recursiva, materializadas en una suma de marcadores de estilo que son estables, prominentes, cuantificables y reconocibles.		
10. Fuentes de la definición: Definición propia basada en la revisión bibliográfica de las fuentes que componen el corpus de referencia de estilometría (cf. 4.1. y Apéndice I).		
11. Nivel de metalengua: general		

- |  |
|--|
| <p>12. Contextos representativos:</p> <ol style="list-style-type: none"><li>“el <i>estilo</i> es inconsciente y deja una especie de huellas estilográficas, y por esta razón, cada quien tiene su <i>estilo</i> propio”</li><li>“la noción de huella lingüística o conjunto de rasgos lingüísticos que caracterizan el <i>estilo</i> de escritura de un autor o una obra y permiten identificarlo como un estilo individual y único que lo diferencia de otros estilos de otros autores u otras obras”</li><li>“un <i>estilo</i> propio puede ser rastreado por medio de métodos cuantitativos (...) (el estilo es) en principio inconsciente, resulta muy difícil de manipular deliberadamente para la mayoría de hablantes o escritores”</li><li>“es que el <i>estilo</i> es algo que nace en el subconsciente, y por esta razón, cada quien tiene su <i>estilo</i> propio, por lo anterior cada autor en su trabajo deja una huella lingüística que lo identifica y diferencia de otros, por lo cual cada <i>estilo</i> es único”</li></ol> |
| <p>13. Fuentes de cada contexto:</p> <ol style="list-style-type: none"><li>Ruiz Cornejo, N. (2019): <i>Análisis estilométrico en Python aplicado a epistolarios de Gabriela Mistral</i>, Trabajo de Fin de Máster, Santiago: Universidad de Santiago de Chile, p. 28.</li><li>Lozano Zahonero, M. (2020): “Una nueva visión de la supuesta influencia de Madame Bovary en La Regenta a través de la estilometría y el análisis de sentimientos basados en lenguaje R”, <i>Orillas: rivista d'ispanistica</i>, 9, p. 578.</li><li>Navarro Peñarrubia, P. (2021): “Estilometría con fines geolingüísticos aplicada al corpus COSER”, <i>Revista de Humanidades Digitales</i>, 6, p. 23.</li><li>Ruiz Cornejo, N. y Arriagada-Benítez, M. (2022): “Análisis estilométrico en Python aplicado a epistolarios de la premio nobel de literatura Gabriela Mistral”, <i>Ingeniare. Revista chilena de ingeniería</i>, 30(2), p. 230.</li></ol>   |
| <p>14. Equivalencias en otras lenguas:</p> <ol style="list-style-type: none"><li><i>style</i> (inglés)</li><li><i>style</i> (francés)</li><li><i>Stil</i> (alemán)</li></ol>   |
| <p>15. Fuentes de las equivalencias:</p> <ol style="list-style-type: none"><li>Wales, K. (2014): <i>A dictionary of stylistics</i>, London: Routledge, p. 397.</li><li>Leclercq, E. (2020) : “La stylométrie comme outil pour la recherche de l’élaboration des chartes médiévales. Le cas de Cambrai au douzième siècle (1131-1200)”, <i>Bulletin du centre d'études médiévales d'Auxerre</i>, BUCEMA, 24(2), p. 2. Disponible en: <a href="http://journals.openedition.org/cem/17772">http://journals.openedition.org/cem/17772</a> (Fecha de consulta: 10/10/2022).</li><li>Ulrich, W. (1972): <i>Wörterbuch linguistische Grundbegriffe</i>, Kiel: Verlag Ferdinand Hirt, p. 150.</li></ol>  |
| <p>16. Fuentes documentales: Corpus de referencia de estilometría (cf. 4.1) y diccionarios especializados en estilística (cf. Registro 15).</p>  |
| <p>17. Bibliografía recomendada: Corpus de referencia de estilometría (cf. 4.1)</p>  |

**18. Observaciones y notas:**

1. En lo referente a cuestiones metodológicas, se ha confeccionado un corpus especializado monolingüe para posteriormente llevar a cabo un procesamiento computacional del mismo mediante un extractor de contextos definitorios (cf. Registro 12) y sus respectivas fuentes bibliográficas (cf. Registro 13) en lenguaje de programación Python, para así automatizar y agilizar el proceso de elaboración de la ficha.
2. El término se encuentra estandarizado en manuales básicos de lingüística y de estilística, pero no registrado lexicográficamente como uso técnico vinculado a la estilometría, ya que no se encuentran obras lexicográficas específicas sobre esta materia.

Tabla 2. Ficha terminológica de *estilo* como uso técnico vinculado a la estilometría siguiendo el modelo *DOCUTERM* (Casas Gómez, 2020, 2022)

## 6. Conclusiones

Este trabajo planteaba como objetivos principales la presentación de la estilometría como técnica de análisis lingüístico-cuantitativa y la definición del término *estilo* apoyada en la elaboración de una ficha terminológica de su uso técnico vinculado a la estilometría. En un primer lugar, se ha delimitado la estilometría como una subdisciplina de la estilística lingüística en constante desarrollo que aún carece de una terminología consolidada y unificada. Si bien encontramos ocurrencias de términos como *estilometría* y *estilo* en bibliografía científica especializada en estilística o lingüística forense, no encontramos un manual especializado o un glosario que recoja los términos relacionados con la estilometría, por lo que hallamos prioritario la contribución al estudio del léxico especializado de este ámbito.

En términos generales, la revisión teórica que se ha llevado a cabo en el presente trabajo demuestra la multidisciplinariedad del término *estilo*, ya que encontramos una amplia variedad de definiciones, dado su carácter heterogéneo y relacional, no solo en materias ajena a la lingüística, sino en distintas ramas o aplicaciones de la misma (i.e. lingüoestilística, lingüística forense, etc.). Para alcanzar una definición fundamentada en los usos reales de *estilo* como uso técnico vinculado a la estilometría, se ha llevado a cabo de forma complementaria una revisión de la bibliografía científica y una búsqueda de contextos definitorios basada en el análisis de un corpus especializado que cuenta con publicaciones científicas sobre la materia en lengua española. El proceso de diseño de este corpus demuestra que la estilometría aún es un campo de actuación en desarrollo, ampliamente trabajada en el panorama anglosajón, pero con un reducido número de publicaciones en el ámbito académico hispanoparlante. Pese a su vinculación a la lingüística forense, su naturaleza interdisciplinar ha desembocado en la diversificación de esta subdisciplina de la estilística lingüística hacia otros ámbitos ajenos a la lingüística, como es el caso de los

trabajos de crítica literaria. Las mismas fuentes del corpus utilizado en este trabajo demuestran que los especialistas de este ámbito están interesándose por las posibilidades que el análisis estilométrico puede ofrecer a sus investigaciones, reflejado en la cantidad de publicaciones en revistas académicas especializadas en estudios literarios que cuentan con trabajos que hacen uso de la estilometría.

El trabajo terminográfico basado en corpus y asistido por ordenador facilita, además, cuestiones metodológicas asociadas a la elaboración de una ficha terminológica y permite que los contextos definitorios seleccionados y la definición final sean lo menos subjetivos posible, ya que se fundamentan en usos reales contenidos en un corpus específicamente diseñado para esta labor. En este artículo hemos presentado un método semi-automático para complementar el trabajo terminográfico mediante el uso de herramientas computacionales para agilizar y facilitar cuestiones referentes a la extracción del conocimiento que rodea a los términos. Los contextos definitorios y metalingüísticos que aparecen en la ficha se corresponden en gran parte con lo planteado en la revisión teórica. La ficha del uso técnico de *estilo* vinculado a la estilometría recoge que la huella lingüística o estilística viene dada por el estilo de un autor, reconocible por el uso de una serie de marcadores de estilo seleccionados de manera inconsciente sobre un amplio repertorio lingüístico. El hecho de que el estilo de un autor se descomponga en unos rasgos lingüísticos prominentes deriva en que el estilo pueda ser cuantificable y también contrastivo, ya que la frecuencia en la que un autor hace uso de unos rasgos lingüísticos específicos es lo que lo diferencia de otros autores.

Por último, hemos de resaltar de nuevo la intersección de la ciencia terminológica con distintos campos del saber y la utilidad del trabajo terminográfico para solventar cuestiones de ambigüedad y facilitar, mediante la clarificación terminológica, la comunicación científica entre expertos de un mismo campo, además de dotar a esta reciente subdisciplina de cierto estatus científico al contar con el trabajo enfocado a alcanzar una terminología precisa y unificada. En concreto, es indispensable señalar la utilidad de aportaciones científicas como la patente *DOCUTERM* (Casas Gómez, 2020, 2022) como método conveniente especialmente diseñado para la difusión del conocimiento especializado y la resolución de cuestiones de ambigüedad o dispersión terminológica.

## 7. Referencias

- AKHMANOVA, O. S. (ed.) (1970): *The Principles and Methods of Linguostylistics: A Course of Lectures*, Moskva: MGU.  
ALARCÓN, R., BACH, C. y SIERRA, G. (2007): “Extracción de contextos definitorios en corpus especializados: Hacia la elaboración de una herra-

- mienta de ayuda terminográfica”, *Revista española de lingüística*, 37(1), pp. 215-246.
- ALLEN, R. F. (1984): *A style-statistical study of “Adolphe”. Preceded by indexes and a description of computer programming for language analysis*, Genève/Paris: Slatkine/Champion.
- ARNAUD, E. (1976) : “Quelques recherches de stylistique quantitative”, *Cahiers du monde hispanique et luso-brésilien*, 27, pp. 21-29.
- AUGER, P. y ROUSSEAU, L. (1977) : *Méthodologie de la recherche terminologique*, Québec: Régie de la langue française.
- BAILEY, R. W. y DOLEZEL, L. (1968): *An Annotated Bibliography of Statistical Stylistics*, Berlin: De Gruyter.
- BALDINGER, K. (1970): *Teoría semántica. Hacia una semántica moderna*, Madrid: Alcalá.
- BALLY, C. (1909 [1970]): *Traité de stylistique française*, Genève: Georg & Cie S.A.
- BOTO BRAVO, M. Á. (2017): “Mapa estilométrico de la narrativa de Eduardo Mendoza: Aproximación a un análisis estilístico computacional de textos literarios”, *Epos: Revista de filología*, 33, pp. 99-114. <https://doi.org/10.5944/epos.33.2017.18281>
- CABRÉ, M. T. (1993): *La terminología. Teoría, metodología, aplicaciones*, Barcelona: Antártida/Empúries.
- CARRIÓN PASTOR, M. (2020): “Conocer la lengua a través de los corpus: la herramienta Metool, retos para el análisis de los marcadores discursivos” *Pragmalingüística*, 28, pp. 255-274. <https://doi.org/10.25267/Pragmalinguistica.2020.i28.13>
- CASAS GÓMEZ, M. (2006): “Modelos representativos de documentación terminográfica y su aplicación a la terminología lingüística”, *Revista de lingüística y lenguas aplicadas*, 1, pp. 25-36.
- CASAS GÓMEZ, M. (2009): “Contenidos y tendencias de la semántica “tradicional” como etapa historiográfica”, García Martín, J. M. (dir.), Bastardín Candón, T. y Rivas Zancarrón, M. (eds.), *Estudios de historiografía lingüística. Actas del VI Congreso Internacional de la Sociedad Española de Historiografía Lingüística en la senda de 1812: Las ideas y realidades lingüísticas en los siglos XVIII y XIX (Cádiz, del 6 al 9 de noviembre de 2007)*, Cádiz: Universidad de Cádiz, pp. 113-150.
- CASAS GÓMEZ, M. (2015): “Semántica léxica, terminología y terminografía: a propósito de los usos terminológicos relacionados con los contenidos y perspectivas actuales de la semántica léxica”, Santana Marrero J., León-Castro Gómez M., Zerva A. (coords.), *La variación en el español actual: estudios dedicados al profesor Pedro Carbonero*, Sevilla: Universidad de Sevilla, pp. 91-110.
- CASAS GÓMEZ, M. (2020): *DOCUTERM: Modelo de documentación de usos terminológicos*, Patente registrada con número 202099908175522, número de expediente: CA-00256-2020 y número de asiento registral 04/2020/4177.
- CASAS GÓMEZ, M. (2021): “La semántica de Kurt Baldinger”, Escavy Zamora, R., Hernández Sánchez, E., Sánchez Manzanares, M. C. (eds.), *La semántica de Kurt Baldinger en la perspectiva de la lingüística actual*, Murcia: Universidad de Murcia, pp. 59-105.
- CASAS GÓMEZ, M. (2022): “DOCUTERM: Modelo de documentación de usos terminológicos”, *Revista de Lingüística y Lenguas Aplicadas*, 17, pp. 11-27. <https://doi.org/10.4995/rlyla.2022.16249>
- COULTHARD, M., JOHNSON, A., y WRIGHT, D. (2016): *An introduction to forensic linguistics: Language in evidence*, London/New York: Routledge.

- CRESPO MIGUEL, M. (2020): "Lingüística digital: revisión de su estado actual y retos en el Instituto Universitario de Investigación en Lingüística Aplicada de la Universidad de Cádiz", *Pragmalingüística*, 28, pp. 148-165. <https://doi.org/10.25267/Pragmalinguistica.2020.i28.08>
- CRESPO MIGUEL, M. (2021): "Las humanidades digitales en el contexto de la lingüística forense: aplicación a sentencias del Consejo General del Poder Judicial", *Derecho y lingüística: entre la concepción teórica y los ciudadanos*, 92, pp. 55-70.
- CRYSTAL, D. (1997): *The Cambridge Encyclopedia of Language*, Cambridge: Cambridge University Press.
- DAELEMANS, W. (2013): "Explanation in computational stylometry", Gelbukh, A. (ed.), *International Conference on Intelligent Text Processing and Computational Linguistics*, Berlin/Heidelberg: Springer, pp. 451-462. [https://doi.org/10.1007/978-3-642-37256-8\\_37](https://doi.org/10.1007/978-3-642-37256-8_37)
- DÍAZ NEGRILLO, A., y DOMÍNGUEZ FERNÁNDEZ, J. (2010): "Corpus terminológicos", Roldán Vendrell, M. (ed.), *Bases para la terminología multilingüe del aceite de oliva, olive oil*, Granada: Comares, pp. 37-50.
- DUBUC, R. (1999): *Manual práctico de terminología. Versión traducida de Ileana Cabrera*. 3. ed. corregida y actualizada, Providencia: Unión Latina, RiL editores.
- DUGAST, D. (1980): *La statistique lexicale*, Genève: Slatkine.
- EDER, M. y KESTEMONT, M. (2016): "Stylometry with R: A Package for Computational Text Analysis", *The R Journal*, 8(1), pp. 107-121. <http://dx.doi.org/10.32614/RJ-2016-007>
- FERNÁNDEZ TRILLO, J. J. (2015): *La problemática autoría de Carlos Castaneda: Un estudio estilométrico de candidato único*. Tesis Doctoral. Madrid: Universidad Autónoma de Madrid.
- FUSTER MÁRQUEZ, M. (2022): "Phraseology and Style in Margaret Jull Costa's English Translation of Marias' Corazón Tan Blanco (1992)", *Pragmalingüística*, 30, pp. 143-162. <https://doi.org/10.25267/Pragmalinguistica.2022.i30.07>
- GRIEVE, J. (2007): "Quantitative Authorship Attribution: An Evaluation of Techniques", *Literary and Linguistic Computing*, 22(3), pp. 251-270. <https://doi.org/10.1093/lrc/fqm020>
- HERNÁNDEZ LORENZO, L. (2022): "Stylistic Change in Early Modern Spanish Poetry Through Network Analysis (with an Especial Focus on Fernando de Herrera's Role)", *Neophilologus*, 106, pp. 397-417. <https://doi.org/10.1007/s11061-021-09717-2>
- HOLMES, D. I. (1998): "The Evolution of Stylometry in Humanities Scholarship", *Literary and linguistic computing*, 13(3), pp. 111-117. <https://doi.org/10.1093/lrc/13.3.111>
- HORTON, T. B. (1987): *The Effectiveness of the Stylometry of Function words in Discriminating between Shakespeare and Fletcher*. Tesis Doctoral. Edinburgh: University of Edinburgh.
- JEFFRIES, L., y MCINTYRE, D. (2010): *Stylistics*, Cambridge: Cambridge University Press.
- JOCKERS, M. L. (2013): *Macroanalysis: Digital methods and literary history*, Urbana: University of Illinois Press.
- KOCK, J., de GEENS, D. (col.), DELBECQUE, N. (col.), GÓMEZ MOLINA, C. (col.) y KEYSER, R. (col.) (1983): *Elementos para una estilística computacional*, Madrid: Coloquio.
- LOZANO ZAHONERO, M. (2020): "Una nueva visión de la supuesta influencia de Madame Bovary en *La Regenta* a través de la estilometría y el análisis

- de sentimientos basados en lenguaje R”, *Orillas: rivista d'ispanistica*, 9, pp. 573-607.
- LUTOSLAWSKI, W. (1898): “Principes de stylométrie appliqués à la chronologie des œuvres de Platon”, *Revue des études grecques*, 11(41), pp. 61-81.
- MÉNARD, N. (1983): *Mesure de la richesse lexicale: théorie et vérifications expérimentales*, Genève/Paris: Slatkine/Champion.
- MORETTI, F. (2000): “Conjectures on world literature”, *New left review*, 1, p. 54.
- MORETTI, F. (2013): *Distant Reading*, London/New York: Verso Books.
- NEAL, T., SUNDARARAJAN, K., FATIMA, A., YAN, Y., XIANG, Y. y WOODARD, D. (2017): “Surveying stylometry techniques and applications”, *ACM Computing Surveys (CSUR)*, 50(6), pp. 1-36. <https://doi.org/10.1145/3132039>
- OLSSON, J. (2008): *Forensic Linguistics: Second Edition*, New York: Continuum International Publishing Group.
- ORGÓNOVÁ, O. (2019): “Aspects of contemporary trends in linguostylistics and in Slovak linguostylistics”, *SKASE Journal of Theoretical Linguistics*, 16(3), pp. 80-89.
- PÉREZ HERNÁNDEZ, M. C. (2002): “Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento”, *Estudios de lingüística del español*, 18. Disponible en: <http://elies.rediris.es/elies18/index.html> (Fecha de consulta: 11/09/2022).
- QUERALT, S. (2020): “El uso de recursos tecnológicos en lingüística forense”, *Pragmalingüística*, 28, pp. 212-237. <https://doi.org/10.25267/Pragmalinguistica.2020.i28.11>
- RAMÍREZ SALADO, M. (2019): *Terminología y lingüística forense: usos terminológicos relacionados con los ámbitos de actuación de la lingüística forense y su interfaz con otras disciplinas*. Tesis Doctoral. Universidad de Cádiz.
- RIFFATERRE, M. (1971): *Essais de stylistique structurale*, Paris: Flammarion.
- SANTAMARÍA PÉREZ, I. (2009): *La terminología: definición, funciones y aplicaciones*, Alicante: Repositorio de la Universidad de Alicante. Disponible en: [http://rua.ua.es/dspace/bitstream/10045/12770/8/La\\_terminología.pdf](http://rua.ua.es/dspace/bitstream/10045/12770/8/La_terminología.pdf) (Fecha de consulta: 11/09/2022).
- SAUSSURE, F. de (1916 [1922]): *Cours de Linguistique Générale. Publié par Ch. Bally et A. Sechehaye avec la collaboration de A. Riedlinger*, 2<sup>a</sup> ed., Paris: Payot.
- SHORT, M. (2018): *Exploring the language of poems, plays and prose*, London / New York: Routledge.
- STAMATATOS, E. (2009): “A survey of modern authorship attribution methods”, *Journal of the American Society for Information Science and Technology*, 60(3), pp. 538-556. <https://doi.org/10.1002/asi.21001>
- ULLMANN, S. (1964): *Language and style*, Oxford: Basil Blackwell.
- ULLMANN, S. (1973): *Meaning and style. Collected Papers*, Oxford: Basil Blackwell.
- VEGA GARCÍA-LUENGOS, G. (2021): “Las comedias de Lope de Vega: confirmaciones de autoría y nuevas atribuciones desde la estilometría (I)”, *Talía. Revista de estudios teatrales*, 3, pp. 91-108. <https://doi.org/10.5209/tret.74625>
- WALES, K. (2014a): *A dictionary of stylistics*, London: Routledge.
- WALES, K. (2014b): “The stylistic tool-kit: Methods and sub-disciplines”, Stockwell, P. y Whiteley, S. (eds.), *The Cambridge handbook of stylistics*, Cambridge: Cambridge University Press, pp. 32-45.

WEIGER, J. G. (1980): "On the Application of Stylostatistics to the Analysis of the Comedia", *Bulletin of the Comediantes*, 32(1), pp. 63-73.

## APÉNDICE I

### *Artículos científicos de revista*

- CELMA VALERO, M. P. y RUIZ URBÓN, C. (2021): "Una aproximación a la escritura de Miguel Delibes desde la estilometría", *Tonos Digital*, 41, pp. 1-13. Disponible en: <http://hdl.handle.net/10201/111055> (Fecha de consulta: 15/09/2022).
- DEMATTÈ, C. (2019): "Una nueva comedia en colaboración entre ¿Calderón?, Rojas Zorrilla y Montalbán: 'Empezar a ser amigos' a la luz del análisis estilométrico", *Rilce: Revista de Filología Hispánica*, 35(3), pp. 852-874. <https://doi.org/10.15581/008.35.3.852-74>
- FRADEJAS RUEDA, J. M. (2016): "El análisis estilométrico aplicado a la literatura española: las novelas policíacas e históricas", *Carácteres: estudios culturales y críticos de la esfera digital*, 5(2), pp. 196-245. <http://dx.doi.org/10.17613/M6J66F>
- GÓMEZ CABALLERO, I. (2021): "Estudio estilométrico del teatro latino: a vueltas con Octavia y Hercules Oetaeus de Séneca", *Cuadernos de Filología Clásica. Estudios Latinos*, 41(1), pp. 57-74. <https://doi.org/10.5209/cfcl.77308>
- GUTIÉRREZ CONSTANTINO, F. (2020): "Evaluación bolsas de palabras como característica estilométrica para atribución de autoría", *Revista Tecnología Digital*, 10(1), pp. 51-64.
- HERNANDEZ LORENZO, L. (2019): "Poesía áurea, estilometría y fiabilidad: métodos supervisados de atribución de autoría atendiendo al tamaño de las muestras", *Carácteres: estudios culturales y críticos de la esfera digital*, 8(1), pp. 189-228.
- LARUMBE, M. C., y CUÉLLAR, Á. (2021): "Discernir entre original y refundición en el teatro del Siglo de Oro a través de la estilometría: el caso de El mejor amigo, el muerto", *Talía: Revista de estudios teatrales*, 3, pp. 59-69. <https://dx.doi.org/10.5209/tret.74021>
- LOZANO ZAHONERO, M. (2020): "Una nueva visión de la supuesta influencia de Madame Bovary en La Regenta a través de la estilometría y el análisis de sentimientos basados en lenguaje R", *Orillas: rivista d'ispanistica*, 9, pp. 573-607.
- NAVARRO PEÑARRUBIA, P. (2021): "Estilometría con fines geolinguísticos aplicada al corpus COSER", *Revista de Humanidades Digitales*, 6, pp. 22-42. <https://doi.org/10.5944/rhd.vol.6.2021.30870>
- NUESCH, E. y RODRIGUES, A. (2021): "Análisis estilométrico de tres obras de Roxana Miranda Rupailaf", *Revista Interfaces*, 12(2), pp. 242-250. <http://dx.doi.org/10.5935/2179-0027.20210040>
- RIOFRÍO, C. E., SÁNCHEZ, W. F., TORRES, A. A., y MORÁN, V. M. (2022): "Identificación de ideología política mediante un modelo Transformer para estilometría y Clasificación por votos en Machine Learning", *Polo del Conocimiento*, 7(9), pp. 1457-1474. <http://dx.doi.org/10.23857/pc.v7i9.4642>
- RODRÍGUEZ LÓPEZ-VÁZQUEZ, A. (2018): "Las dos partes del Lazarillo y su autor. Un estudio de estilometría", *Janus: estudios sobre el Siglo de Oro*, 7, pp. 55-91.

- RODRÍGUEZ LÓPEZ-VÁZQUEZ, A. (2019): “Cervantes y helecho de Procusto: notas críticas al uso de la Estilometría en obras de atribución dudosa y en obras apócrifas”, *eHumanista: Journal of Iberian Studies*, 41, pp. 193-201.
- RUIZ CORNEJO, N. y ARRIAGADA-BENÍTEZ, M. (2022): “Análisis estilométrico en Python aplicado a epistolarios de la premio nobel de literatura Gabriela Mistral”, *Ingeniare. Revista chilena de ingeniería*, 30(2), pp. 229-238. <http://dx.doi.org/10.4067/S0718-33052022000200229>
- TELLO, J. C., y SOLER, J. C. (2018): “La conquista de Jerusalén ¿de Cervantes? Análisis estilométrico sobre autoría en el teatro del Siglo de Oro español”, *Digital Humanities Quarterly*, 12(1). Disponible en: <https://www.proquest.com/scholarly-journals/la-conquista-de-jerusalen-cervantes-analisis/docview/2555195012/se-2> (Fecha de consulta: 15/09/2022).
- VEGA GARCÍA-LUENGOS, G. (2021): “Las comedias de Lope de Vega: confirmaciones de autoría y nuevas atribuciones desde la estilometría (I)”, *Talia. Revista de estudios teatrales*, 3, pp. 91-108. <https://doi.org/10.5209/tret.74625>
- VÉLEZ, J. I., y MARMOLEJO-RAMOS, F. (2016): “Los secretos de Cien Años de Soledad: Una aproximación estilométrica para la investigación en Psicolingüística”, *Revista Colombiana de Psicología*, 25(2), pp. 265-288. <https://doi.org/10.15446/rcp.v25n2.50742>

### *Capítulos de libro*

- BLASCO PASCUAL, F. J. (2016): “Avellaneda desde la estilometría”, Ruiz Pérez, P. (ed.), *Cervantes: los viajes y los días*, Madrid: Sial Ediciones, pp. 97-112.
- BLASCO PASCUAL, F. J. (2019): “Atribuciones cervantinas desde la estilometría: el entremés de Los mirones”, Laín Corona, G. y Santiago Nogales, R. (eds.), *Cartografía literaria en homenaje al profesor José Romera Castillo*, Madrid: Visor Libros, pp. 151-168. Disponible en: <http://uvadoc.uva.es/handle/10324/37760> (Fecha de consulta: 15/09/22).
- BLASCO PASCUAL, F. J. (2018): “Introducción a la estilometría: Algunos ejemplos de técnicas y protocolos”, Álvarez Ramos, E. y Blasco Pascual, J. (eds.), *Humanidades digitales: Retos, recursos y nuevas propuestas*, Valladolid: Universidad de Valladolid: Agilice Digital, pp. 23-34.
- FRADEJAS RUEDA, J. M. (2019): “Estilometría y la Edad Media castellana”, Rißler-Pipka, N. (ed.), *Romanische Studien. Theorien von Autorschaft und Stil in Bewegung: Stilistik und Stilometrie in der Romania*, München: Akademischen Verlagsgemeinschaft, pp. 49-74.
- HERNÁNDEZ LORENZO, L. (2019): “Fernando de Herrera y la autoría de Versos: un primer acercamiento al drama textual desde la Estilometría”, Rißler-Pipka, N. (ed.), *Romanische Studien. Theorien von Autorschaft und Stil in Bewegung: Stilistik und Stilometrie in der Romania*, München: Akademischen Verlagsgemeinschaft, pp. 75-90.

### *Actas de Congresos*

- BIA PLATAS, A. (2016): “Estilometría computacional, algunas experiencias en el marco del proyecto TRACE”, del Río Riande, G., Calarco, G., Striker, G., de León, R. (eds.), *Humanidades Digitales: Construcciones locales en contextos globales. Actas del I Congreso Internacional de la Asociación*

- Argentina de Humanidades Digitales-AAHD*, Buenos Aires: Universidad de Buenos Aires, pp. 39-46.
- HERNÁNDEZ LORENZO, L., y DE ORO, P. D. S. (2017): “Estilística computacional y estilometría aplicadas a la poesía en español: el caso de Fernando de Herrera”, *III Congreso de la Sociedad Internacional Humanidades Digitales Hispánicas Sociedades, políticas, saberes (Libro de resúmenes)*, 18(20), p. 254.

*Tesis doctorales y otros trabajos académicos*

- FERNÁNDEZ TRILLO, J. J. (2015): *La problemática autoría de Carlos Castañeda: Un estudio estilométrico de candidato único*. Tesis Doctoral. Universidad Autónoma de Madrid.
- HERNÁNDEZ LORENZO, L. (2020): *Los textos poéticos de Fernando de Herrera: aproximaciones desde la estilística de Corpus y la Estilometría*. Tesis Doctoral. Universidad de Sevilla.
- MALDONADO GALIANO, J. R. (2015): *Aplicación de estilometría para la atribución autorías en e-mails y documentos informáticos*. Trabajo de Fin de Grado. Universidad de San Francisco de Quito.
- MOYANO MORENO, I. (2021): *Acercamiento a las Humanidades Digitales y la Estilometría: Propuesta de análisis estilométrico con Stylo (R) sobre un corpus de literatura anglófona realista y modernista*. Trabajo de Fin de Máster. Universidad de Cádiz. Disponible en: <http://hdl.handle.net/10498/25113>
- RUIZ CORNEJO, N. (2019): *Análisis estilométrico en Python aplicado a epistolarios de Gabriela Mistral*. Trabajo de Fin de Máster. Universidad de Santiago de Chile. Disponible en: <https://repositorio.uc.cl/handle/11534/23001> (Fecha de consulta: 30/09/2022).
- VALENCIA UY, A. (2018): *Francisco Laguna desde la estilometría y la crítica genética*. Trabajo de Fin de Máster. Universidad de Valladolid.