

CRONOLOGÍA Y ESTILOMETRÍA: DATACIÓN AUTOMÁTICA
DE COMEDIAS DE LOPE DE VEGA*

ÁLVARO CUÉLLAR (Universität Wien)

CITA RECOMENDADA: Álvaro Cuéllar, «Cronología y estilometría: datación automática de comedias de Lope de Vega», *Anuario Lope de Vega. Texto, literatura, cultura*, XXIX (2023), pp. 97-130.

DOI: <<https://doi.org/10.5565/rev/anuariolopedevega.483>>

Fecha de recepción: 4 de julio de 2022 / Fecha de aceptación: 6 de septiembre de 2022

RESUMEN

La datación supone una de las cuestiones con más interrogantes en el estudio del teatro aurisecular español. La estilometría, técnica informática basada en la comparación de textos en función de sus usos escriturales y que está utilizándose con aparente acierto en los últimos años para arrojar luz sobre el problema de la autoría, puede servir también para hallar el momento de escritura de un texto. En este artículo presento un primer acercamiento para comprobar tal extremo. Para ello, aplico un análisis estilométrico a un corpus controlado de autógrafos de Lope de Vega con fecha de composición segura haciendo uso de tres elementos base: palabras, lemas y categorías gramaticales. Los resultados parecen mostrar una correlación entre estilo y datación, obteniéndose errores absolutos medios en el entorno de ± 5 años, franja estrecha y prometedora si tenemos en cuenta la dilatada producción de Lope de Vega. Aplico, en segundo lugar, la prueba con palabras a un conjunto más amplio de comedias lopescas de datación aproximada, con resultados positivos en el entorno de ± 4 años de error absoluto medio.

PALABRAS CLAVE: Datación; Teatro; Estilometría; Siglo de Oro; Lope de Vega.

ABSTRACT

Finding the accurate date of texts is one of the most puzzling challenges in the study of Spanish aurisecular theater. Stylometry, the computer technique centered on the comparison of texts based on their writing, which has been successfully used in recent years to solve authorship questions, can also serve to shed light on this issue. In this article I present a first approach to verify such idea. I apply stylometric analysis to a controlled corpus of Lopean autographs using three basic elements: words, lemmas and parts of speech. The results seem to show a correlation between style and dating, obtaining mean absolute errors around ± 5 years, a fairly narrow range if we take into account the long production of Lope de Vega. Secondly, I apply the test with words to a broader set of Lopean plays of approximate date with positive results, close to mean absolute errors of ± 4 years.

KEYWORDS: Dating; Theater; Stylometry; Spanish Golden Age; Lope de Vega.

* Este artículo no podría haber llegado a término sin la atenta ayuda de Germán Vega García-Luengos (Universidad de Valladolid). La investigación está amparada por el proyecto *Sound and Meaning in Spanish Golden Age Literature* (Universität Wien) (FWF Austrian Science Fund 32563). Los corpus de frecuencias que permiten replicar este experimento, así como ampliarlo y mejorarlo, se encuentran en el repositorio <https://github.com/AlvaroCuellar> o mediante petición al autor.

Este es, por lo tanto, un intento de considerar el arte métrico de un escritor como un crecimiento natural, y estudiarlo del mismo modo que un botánico estudia el desarrollo progresivo de una planta. Un hombre no se da cuenta de los cambios que se operan en su rostro año tras año y, sin embargo, ocurren en la realidad. Lope tampoco se dio cuenta de que obedecía a una ley de crecimiento y de que escribía sometido a ella.

Sylvanus Griswold Morley y Courtney Bruerton
Cronología de las comedias de Lope de Vega (1968)

1. INTRODUCCIÓN

La estilometría, conjunto de técnicas que permite comparar estadísticamente los textos por sus usos léxicos, está demostrando en los últimos años su capacidad para relacionar obras del teatro aurisecular hispano en función de su autoría. Las frecuencias de uso de las palabras empleadas por los distintos autores parecen conformar una huella escritural particular y pueden ayudarnos a inferir conclusiones de autoría a través de relaciones de proximidad o lejanía entre los textos.

Estas relaciones, tan fructíferas para las atribuciones autoriales, —como se expondrá en el siguiente apartado—, parecen, también, trascender la autoría y ayudarnos a abordar otros problemas del universo literario, como el estudio de los géneros (Calvo Tello 2021), la procedencia geográfica (Peñarrubia Navarro 2021), la genealogía (O’Sullivan 2021), las refundiciones (Campión Larumbe y Cuéllar 2021), etc.

Además de arrojar luz al respecto de estas cuestiones, la estilometría también podría, a priori, servir para la datación de textos teatrales del periodo aurisecular. Se trata de una vía de exploración todavía en ciernes, sobre la que específicamente poco o nada se ha trabajado.¹ La estilometría, como puede apreciarse en la bibliografía, es una técnica reciente, sobre la que se está empezando a investigar en los últimos años. Su aplicación al teatro del Siglo de Oro y el intento de

1. Para un acercamiento estilométrico a la evolución del estilo a través del tiempo en la obra de Miguel Delibes, véase Celma Valero y Ruiz Urbón [2021].

datación supone un paso arriesgado, pero también ilusionante. Espero, por tanto, que otros trabajos vengan pronto a sumarse a este y a refinar los procedimientos aquí descritos.

La datación, esto es, el intento por encontrar la fecha de composición de una pieza literaria, supone uno de los retos más desafiantes para el investigador del teatro del Siglo de Oro. Las obras dramáticas, que se cuentan por miles entre los siglos XVI y XVII, debido a la popularidad absoluta que alcanzó el género entre todos los estratos de la sociedad barroca, aparecen, en muchos casos, sin fecha alguna y diseminadas en multitud de testimonios, tanto impresos como manuscritos. Si bien es cierto que los impresos en volúmenes o partes suelen ofrecer fechas, estas muy raramente corresponden con la de creación de las obras, pues las comedias son dadas a la imprenta habitualmente mucho después de su representación e incluso después de la muerte del dramaturgo.

De Lope de Vega, debido al gran número de obras que fue capaz de escribir, conservamos unas 360 comedias —esta es la cifra propuesta en el reciente trabajo de Vega García-Luengos [2021]—, aunque es de esperar que sigan produciéndose altas, a la vista de investigaciones como las de García-Reidy [2019] o Madroñal [2021]. La producción dramática de Lope de Vega supone el campo perfecto de estudio para la problemática de la datación y para su abordaje desde procesos informáticos automáticos de estilometría. Estos cientos de obras recorren la vida de Lope (1562-1635) de forma homogénea: fue prolijo prácticamente en cada década desde que empieza a escribir hasta su muerte, independientemente de su ajetreada vida personal y las distintas actividades a las que se entregaba, literarias o no.

En el famoso estudio de Morley y Bruerton [1968], titulado en español *Cronología de las comedias de Lope de Vega. Con un examen de las atribuciones dudosas, basado todo ello en un estudio de versificación estrófica*, se asienta un pilar fundamental para los trabajos de datación en el teatro del Siglo de Oro y, específicamente, para la producción lopesca. Morley y Bruerton fueron pioneros en aplicar argumentos de carácter matemático-estadístico a un problema histórico-filológico. En su ya clásica obra dan cuenta de la producción conocida de Lope de Vega, hacen un repaso de las obras con datación segura y proponen fechas para las obras no datadas. Para ello, utilizan un sistema basado en el análisis estadístico de la métrica de las obras de Lope y sus proporciones de uso. Así, tienen en cuenta el porcentaje de redondillas, quintillas, décimas, romances, silvas, octavas, sonetos, tercetos, liras, versos

sueltos, y demás metros menos comunes que, según observan los investigadores, Lope utilizaba en distinta proporción a lo largo de su producción dramática.

Al aproximarse a una nueva obra, Morley y Bruerton llevan a cabo un conteo de las proporciones de estos usos métricos y sitúan el texto en una franja de tiempo determinada. Este proceso parece ser bastante certero, pues a día de hoy las tablas generadas por los investigadores siguen siendo guía para las generaciones actuales de estudiosos del teatro del Siglo de Oro. Su análisis permitía, también, detectar obras que por sus usos métricos no encajaban en la producción lopesca y que, con el paso de las décadas y los estudios de nuevos investigadores, se han acabado demostrando como no auténticas de Lope. En definitiva, el método desarrollado por Morley y Bruerton supuso una piedra de toque para el conocimiento de la datación de la producción lopesca. El sistema empleado por estos, sin embargo, se demuestra en ocasiones un tanto opaco al investigador que pretende replicarlo. Resulta divertido incluir aquí el comentario del propio Morley [1968:12] en su prólogo a la versión de 1968, donde dice: «así, en mi escala menor, he olvidado las técnicas exactas del método, en el que una vez estuvimos tan profundamente inmersos, y he tenido que recuperarlas como del vacío». Si es complicado para el propio Morley, imaginemos para el estudioso que decide intentar utilizar el sistema con un nuevo texto. Además, el método propuesto por Morley y Bruerton es de difícil aplicación a la producción de otros dramaturgos u otros periodos, géneros o literaturas, puesto que está directamente basado en los usos métricos de Lope de Vega. En este artículo comprobaremos si el camino de la estilometría puede ofrecer también una datación certera del teatro de Lope de Vega, con lo que conseguiríamos abrir una vía original al estudio cronológico de otros dramaturgos o géneros de interés.

2. ESTILOMETRÍA MÁS ALLÁ DE LA AUTORÍA

Si Morley y Bruerton trabajaban de forma manual a través de la lectura y anotación cuidadosa de los usos métricos de las obras de Lope de Vega, en la actualidad podemos realizar procesos masivos y automáticos de forma mucho menos esforzada. La estilometría, técnica capaz de comparar textos por el uso de sus elementos básicos, como, por ejemplo, palabras, se ha venido desarrollando en los últimos años en el estudio de la literatura y, particularmente, en el teatro del Siglo de Oro.

En numerosos trabajos, como Cerezo Soler y Calvo Tello [2019], Demattè [2019], Rodríguez-Gallego [2019], Ulla Lorenzo, Martínez Carro y Calvo Tello [2020], Vega García-Luengos [2020], Demattè [2021], Marcos Rodríguez [2021], Martínez Carro [2021 y 2022], Cuéllar [en prensa], etc., se ha aplicado la estilometría para estudiar problemas de autoría que afectan al teatro aurisecular. En el proyecto *ETSO: Estilometría aplicada al Teatro del Siglo de Oro* (<<https://etso.es/>>), que tengo la suerte de codirigir junto a Germán Vega [2017-2022] y que cuenta con unos 2800 textos teatrales de 350 dramaturgos diferentes (unos 36 millones de palabras) constituyentes de *CETSO: Corpus de Estilometría aplicada al Teatro del Siglo de Oro* (<https://etso.es/cetso>), en estos momentos (verano de 2022) estamos preparando los informes estilométricos de autoría de todas las obras de las que disponemos y daremos cuenta de ello en el portal web y en futuras publicaciones. A tenor de estos estudios, parece que la medida y comparación de la frecuencia de uso de las palabras, que es el elemento más utilizado para los análisis, puede ayudarnos a discernir autoría en el teatro del Siglo de Oro.

Por ejemplo, en la Tabla 1 podemos ver las frecuencias de 10 palabras comunes en tres obras capitales del teatro del Siglo de Oro de tres dramaturgos distintos: *Peribáñez* de Lope de Vega, *La dama duende* de Calderón de la Barca y *Don Gil de las calzas verdes* de Tirso de Molina.

	amor	dios	mujer	vida	alma	puerta	honor	gente	hermano	cielos
Peribáñez	0,26%	0,22%	0,26%	0,18%	0,07%	0,10%	0,16%	0,16%	0,01%	0,03%
La dama duende	0,14%	0,26%	0,23%	0,19%	0,04%	0,18%	0,11%	0,11%	0,22%	0,12%
Don Gil de las calzas verdes	0,39%	0,20%	0,14%	0,10%	0,23%	0,03%	0,04%	0,04%	0,01%	0,08%

Tabla 1. Comparación de las frecuencias de uso de 10 palabras comunes en 3 obras capitales del teatro aurisecular.

En *Peribáñez* por cada 100 palabras 0,26 son *amor*, 0,22 son *dios* y 0,26 son *mujer*, mientras que en *La dama duende* las proporciones son 0,14, 0,26 y 0,23. Estas pequeñas diferencias entre las frecuencias de las palabras y sus implicaciones para la autoría no pueden ser entendidas por el limitado cerebro humano, sobre todo si tenemos en cuenta que normalmente para que los análisis funcionen y señalen correctamente hacia un autor en el teatro del Siglo de Oro, no son las tenidas en cuenta solo

10 palabras, sino cientos o miles de estas, Cuéllar [en prensa]. Si preparamos un corpus de teatro aurisecular de varios autores, la estilometría es capaz de decirnos cuáles son las obras con un uso léxico más cercano a la de nuestro interés. Estas obras suelen corresponder, en la mayoría de los casos, a las del autor de nuestra obra. Es decir, la estilometría es capaz, con el corpus y parámetros adecuados, de detectar autoría.

En el caso de Lope de Vega, podemos comprobar este extremo a través de la Figura 1 donde se ha procedido a trazar una red estilométrica automática —teniendo en cuenta las 500 palabras más frecuentes (MFW), las palabras no tienen que estar en un porcentaje mínimo de textos (0% culling) y un método para calcular distancias Classic Delta— con todo el corpus CETSO y se han marcado en rojo las obras tradicionalmente atribuidas a Lope sin ninguna problemática asociada. Enseguida apreciamos que la estilometría ha sido capaz de relacionar en general las

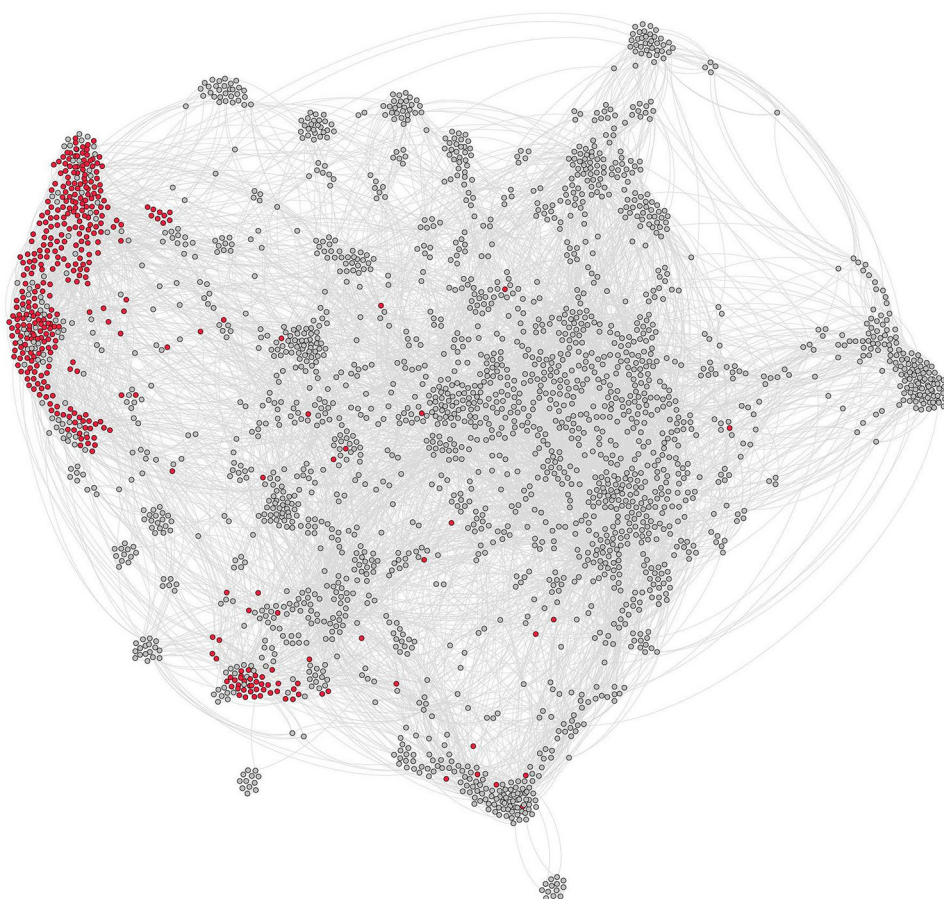


Figura 1. Red estilométrica de 2800 obras de teatro aurisecular. En rojo la producción tradicionalmente atribuida sin dudas a Lope. 500 MFW, 0% culling, Classic Delta.

obras en función del dramaturgo que las ha escrito, aunque, por supuesto, con decenas de casos fronterizos y dudosos que es preciso investigar pormenorizadamente.

Sin embargo, más allá de las cuestiones de autoría, la estilometría puede servirnos también para otros propósitos. Aquí trataré de comprobar si esta técnica es también útil para la datación de obras teatrales a través, en primer lugar, de un corpus controlado de Lope de Vega compuesto por autógrafos fechados. Nuestra hipótesis es, por tanto, la siguiente: el uso de las palabras puede no solo servir para dilucidar autoría, sino que también puede relacionar obras por su fecha de escritura. Una vez superada la señal autorial es posible que afloren otras relaciones de interés, como la del momento de creación del texto.

Una primera impresión de esta hipótesis nos la puede dar la Figura 2, para la que he tomado un conjunto de 296 obras de Lope de Vega de autoría y datación seguras, que será explicado en apartados siguientes y a la que he aplicado los mismos parámetros que en la figura previa. He graduado el color verde en función de la fecha de composición de las obras: cuanto más claro el verde, más tempranas, y cuanto más oscuro, más tardías. Las fechas han sido extraídas principalmente del ya comentado estudio de Morley y Bruerton. En los casos en los que se ofrece una franja para la datación, he tomado el valor medio como año de referencia aproximado.

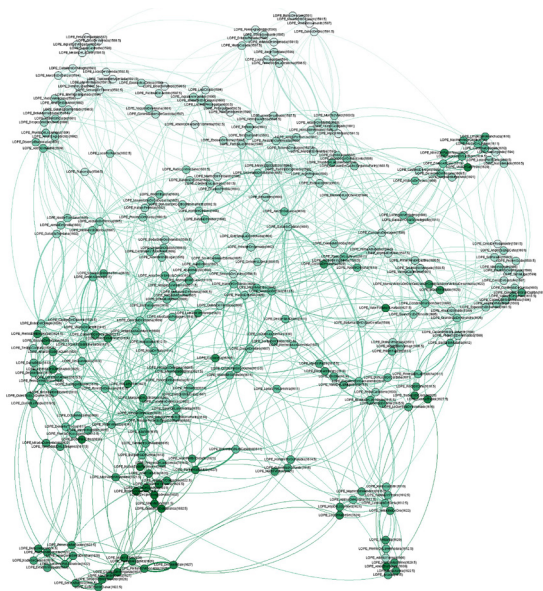


Figura 2. Red estilométrica de 296 obras lopescas. Cuanto más oscuro el verde, más tardía es la obra. 500 MFW, 0% culling, Classic Delta.

Se trata este de un primer acercamiento para comprobar si la estilometría puede relacionar las obras en función de su fecha de composición. En esta red construida automáticamente que aproxima o aleja los textos por sus usos léxicos —en este caso se ha tenido en cuenta el uso de las 500 palabras más frecuentes en el corpus— se puede apreciar que, en efecto, se está produciendo una relación especial dada por la fecha de composición.

Aunque existan muchas irregularidades, en la parte baja de la red podemos apreciar obras más tempranas mientras que en la parte alta aparecen obras más tardías en la producción de Lope. Conviene destacar que otros factores parecen influir con más peso que la datación en la relación entre las obras. Así en la parte derecha encontramos un pequeño conjunto de textos como *La niñez del padre Rojas*, *El nacimiento de Cristo*, *La vida de san Pedro Nolasco*, etc., las cuales comparten unas características temáticas de tipo religioso que sobrepasan la relación por datación que buscábamos. Lo mismo sucede con otros casos en los que se producen relaciones por temática o género dentro del corpus lopesco. Estamos, por tanto, ante una lucha de fuerzas estilométricas² que debemos tener muy en cuenta a la hora de hallar la datación de un texto, puesto que pueden jugar un papel inadecuado en los análisis y debemos ser siempre precavidos al confrontarlas.

En definitiva, parece plausible que, al igual que Morley y Bruerton detectaron una evolución en las formas métricas a lo largo de la producción de Lope, también se pueda apreciar una evolución de otros elementos constituyentes del análisis estilométrico. Para el primer experimento con autógrafos voy a estudiar los que creo son básicos y, a priori, más potentes, en el caso que nos ocupa: palabras, lemas y categorías gramaticales. Es preciso apuntar que la elección de los parámetros para las pruebas estilométricas supone uno de los mayores retos para el investigador. Si aplicamos para el análisis estilométrico la librería Stylo (versión 0.7.4) de Eder, Rybicki y Kestemont [2016] contamos con innumerables opciones distintas de análisis que podríamos variar *ad infinitum*. Podemos modificar las MFW (most frequent words) tenidas en cuenta, el culling (porcentaje de aparición de las palabras en los textos), n-grams (si se trata de elementos simples o combinaciones), método estadístico para medir las distancias (Classic, Manhattan, Canberra, etc.), método

2. Gaston Gilabert y yo estamos en este momento realizando un estudio exhaustivo sobre el peso de las distintas fuerzas estilométricas en Lope de Vega.

clasificadorio (SVM, NSC, k-NN, etc.) por lo que vamos a toparnos con cierta arbitrariedad al elegir estos parámetros. Para este estudio he optado por utilizar 500 MFW (sin eliminar las palabras funcionales), 0% culling, palabras independientes y método Classic Delta, por ser los valores empleados habitualmente en los análisis autoriales de ETSO. Variar estos parámetros no parece tener un impacto demasiado importante en las pruebas, aunque convendría comprobar cuidadosamente este extremo, lo que dejamos para futuras investigaciones.

3. DATACIÓN AUTOMÁTICA DE AUTÓGRAFOS LOPESCOS

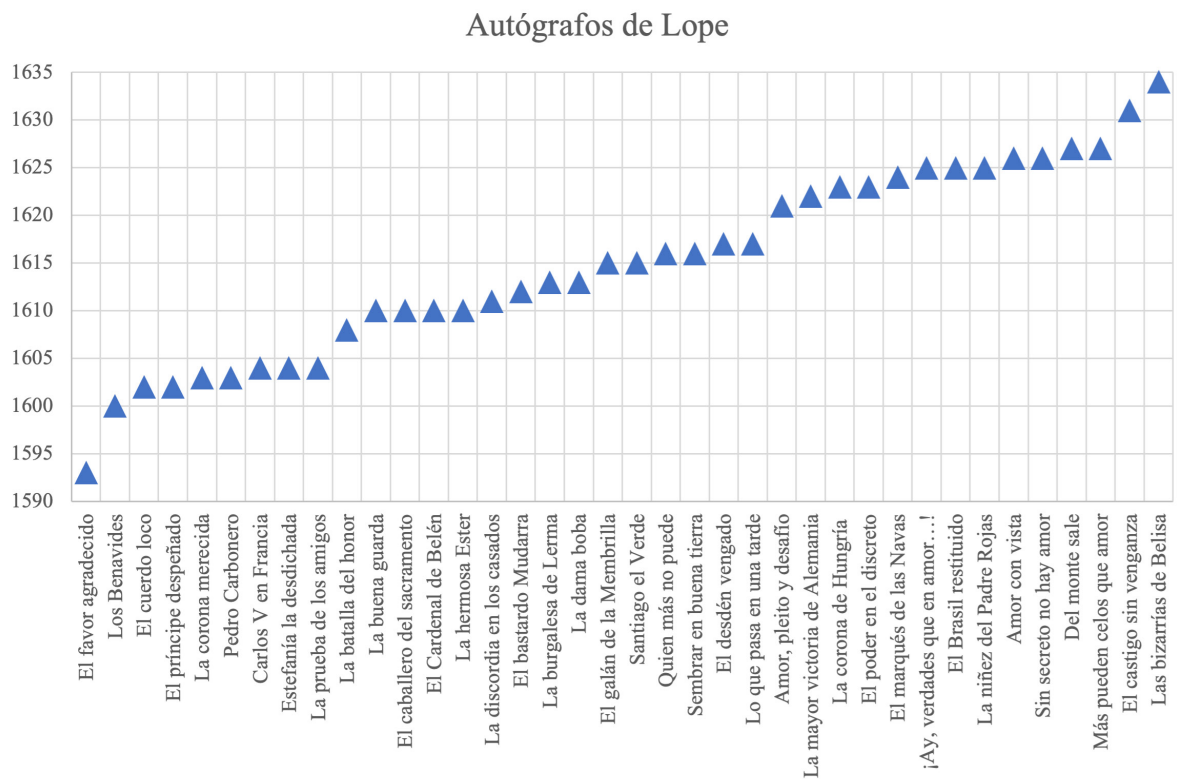
Aunque hemos apuntado la relación que se puede percibir entre datación y uso de palabras con la red del apartado anterior, esto no supone más que una muestra cualitativa, que depende en último término de nuestras impresiones sobre los resultados a través del soporte visual. Es preciso elaborar un experimento con obras controladas, esto es, de las que contemos con una fecha de composición segura en la medida de lo posible. Sobre este corpus de obras seguras realizaremos las pruebas informáticas sabiendo, por tanto, siempre cuál debe ser el resultado. De esta manera podremos calibrar las herramientas y comprobar hasta qué punto son eficaces para nuestro propósito: hallar la fecha de composición de una obra teatral de Lope de Vega.

El corpus para el experimento se corresponde con manuscritos autógrafos de Lope de Vega. Considero este un corpus reducido, controlado y homogéneo, ideal para acometer este primer intento de datación. Presotto [2000] propone en su monográfico un conjunto de 44 autógrafos de Lope de Vega, esto es, testimonios conservados directamente de la pluma del dramaturgo. De estos autógrafos, solo de 38 contamos con la fecha exacta de composición, la cual aparece en el propio documento. Aunque estas fechas salen en principio de la mano de Lope de Vega, conviene reflexionar sobre la idea de que no existe certeza absoluta de que no hayan sido compuestas en un momento y, finalmente, terminadas y fechadas por Lope en otro posterior. Debemos, en todo caso, dar por buenas estas fechas para poder avanzar en el experimento.

En la Gráfica 1 podemos apreciar las 38 obras autógrafas de datación segura. El corpus, que se ha tomado de CETSU y donde pueden consultarse las fuentes

exactas, ha sido homogeneizado y se han eliminado nombres de personajes y acotaciones. El ideal es que todas las obras del corpus procedan de ediciones críticas, pero esto no ha sido siempre posible.

Aquí ya podemos visualizar la homogeneidad en la distribución de los autógrafos lopescos fechados, desde *El favor agradecido* en 1593 hasta *Las bizzarrias de Belisa* en 1634, los cuales nos pueden servir bien para la prueba de datación automática.



Gráfica 1. Autógrafos fechados de Lope de Vega.

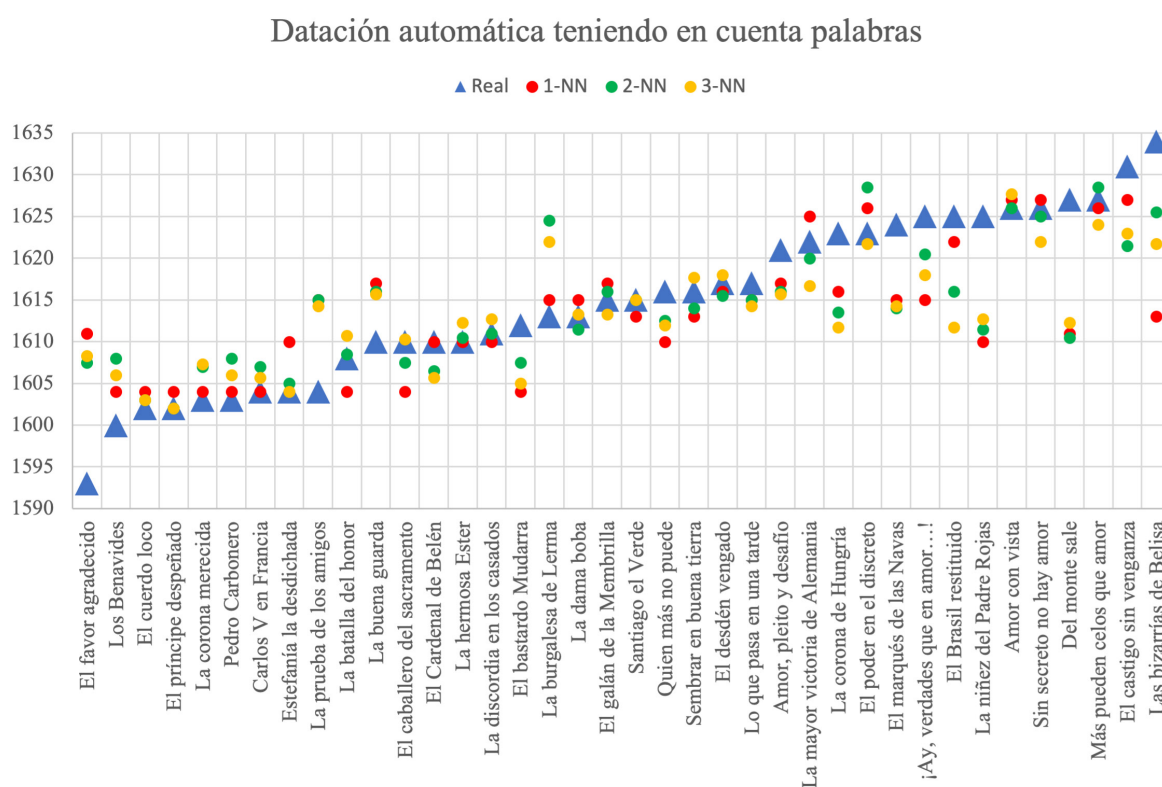
3.1 Datación mediante palabras

Para el primer experimento de datación automática utilizaremos las palabras como los elementos básicos de los textos que se van a comparar. Haremos uso de las 500 palabras más frecuentes en el corpus (*que, de, y, el, a, la, en, no, es, mi, etc.*), una distancia Classic Delta³ (propuesta por Burrows), 0% culling (las palabras no tienen que

3. Para una explicación intuitiva de Delta acudir a Calvo Tello [2016].

estar en un porcentaje determinado de los textos) y usaremos palabras independientes en vez de grupos de palabras. Esta elección de parámetros, como apuntaba antes, es arbitraria, son los parámetros que se emplean en el proyecto ETSO y parecen estar dando buenos resultados para la relación de obras en función de su autoría.

Una vez realizado este análisis para cada texto, obtendremos una lista de las obras más cercanas a cada una por sus usos léxicos. Para intentar datar la obra, tendremos en cuenta el año de producción de las más cercanas, lo que en ciencia estadística se suele denominar procesos de *nearest neighbor*, aunque aquí no se aplica como un complejo *machine learning*, sino que simplemente se apuntará la fecha que corresponde a la obra más cercana (1-NN), la media de las dos obras más cercanas (2-NN) y la media de las tres obras más cercanas (3-NN). Los resultados se pueden observar en la Gráfica 2 en la que con un triángulo negro aparece la datación tradicional de la obra y los puntos indican el intento de datación automática teniendo en cuenta una (rojo), dos (azul) y tres (verde) obras más cercanas.



Gráfica 2. Datación automática del corpus autógrafo de Lope teniendo en cuenta palabras. 500 MFV, 0% culling, Classic Delta.

El error absoluto medio, la media de la diferencia entre el valor real y el valor averiguado, calculado como sigue:

$$\text{EAM} = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

es el siguiente: 1-NN = $\pm 4,9$; 2-NN = $\pm 5,2$; 3-NN = $\pm 5,1$; es decir, hemos podido datar automáticamente las obras con un margen de error medio aproximado de ± 5 años. Se trata de un resultado bastante certero. Como se aprecia en la figura, el acierto no es homogéneo, mientras las obras de la primera mitad de la producción lopesca se datan con un gran éxito, las obras de la segunda mitad presentan una mayor desincronización. Cabe señalar que, evidentemente, la primera y la última obra en ningún caso pueden resultar bien datadas con este método porque no hay otras con su misma fecha y la media de las dos o tres más cercanas no puede nunca arrojar el año real. En todo caso, parece plausible afirmar la existencia de relación estilométrica por la fecha de composición de la obra.

3.2 *Datación mediante lemas*

La estilometría puede utilizar distintos elementos para la comparación estadística. En el caso anterior eran analizadas las palabras más frecuentes. Es posible que si cambiamos estos componentes los resultados varíen y se ajusten más a nuestros propósitos. Sería, en principio, esperable que transformar el corpus en lemas en vez de usar palabras tuviera repercusión en los resultados. Podría suceder que al comparar obras estén pesando factores como el tiempo en el que suelen hablar los personajes (mayor preminencia de pasado o de futuro, por ejemplo) o el género de estos (mayor preminencia de personajes masculinos o femeninos, por ejemplo). Si sustituimos las palabras por sus correspondientes lemas, es posible que la datación automática mejore. Para ello se ha tomado *Freeling* de Padró y Stanilovsky [2012] en su versión 4.0, que es el lematizador más popular para el español, y lo he aplicado al corpus de autógrafos de Lope de Vega, transformando todas las palabras en sus correspondientes lemas.

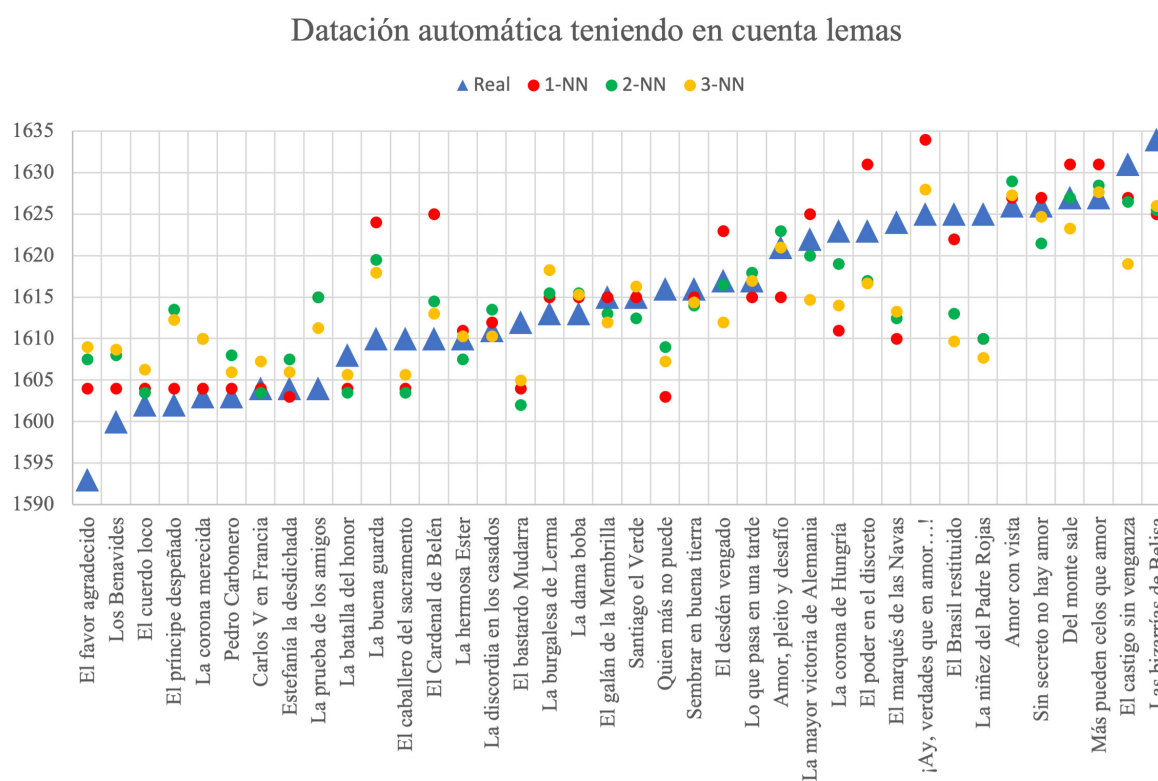
Así, versos del tipo:

orillas hermosas
que el cristal cubrís
tened que me muero
lástima de mí

Quedan transformados en:

orilla hermoso
que el cristal cubrir
tener que me morir
lástima de mí

Se aplica el mismo proceso que en el apartado anterior, con los mismos parámetros (500 MFW, 0% culling, palabras independientes, Classic Delta) obteniéndose los resultados reflejados en la Gráfica 3:



Gráfica 3. Datación automática del corpus autógrafo de Lope teniendo en cuenta lemas. 500 MFW, 0% culling, Classic Delta.

En este caso tenemos un error absoluto medio de 1-NN = $\pm 4,9$; 2-NN = $\pm 5,2$; 3-NN = $\pm 5,2$. Los resultados, por tanto, parecen muy similares a los obtenidos con palabras, apenas hay diferencias significativas en los errores absolutos medios.

3.3 *Datación mediante categorías gramaticales*

En último lugar, podemos acometer un cambio en el corpus mucho más profundo que el de utilizar lemas: es posible transformar el corpus en categorías gramaticales. En vez de palabras como elementos constituyentes de los textos, ahora contamos con las categorías gramaticales correspondientes. De esta manera versos como los siguientes:

tarde pienso que venimos
sin misa nos quedaremos
la intención ofreceremos
culpa de tardar tuvimos

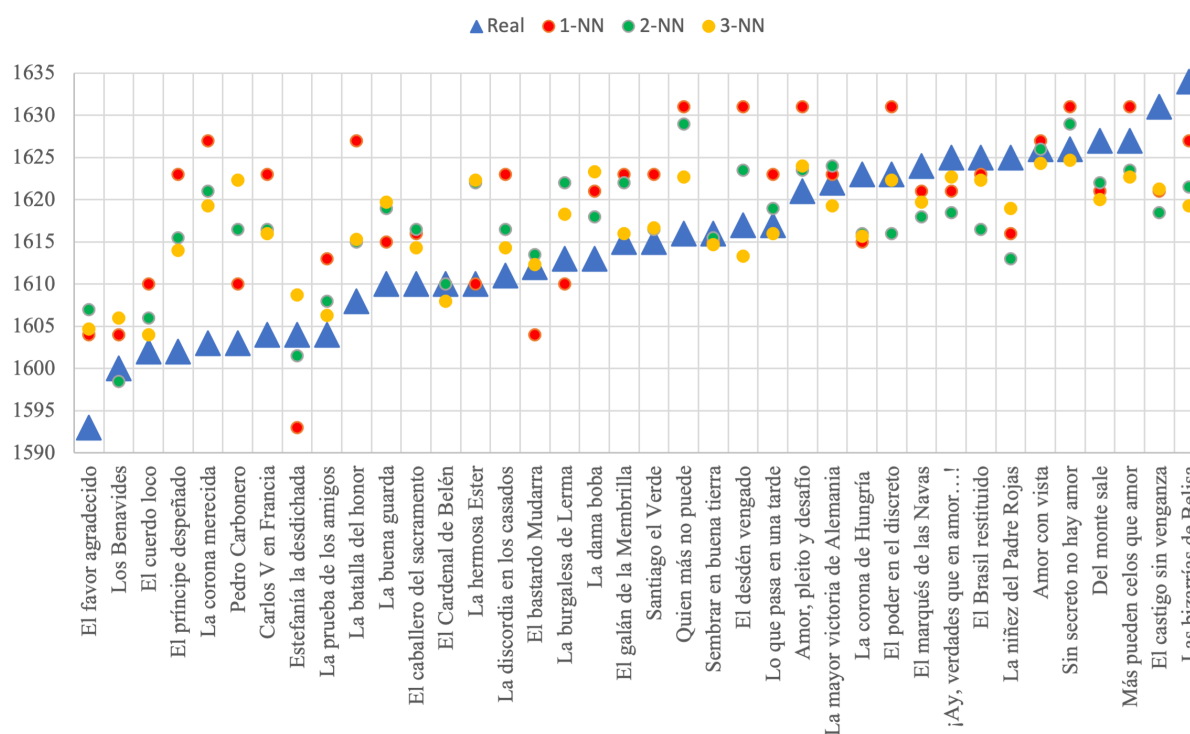
Quedan transformados en:

RG VMIP1S0 CS VMIP1P0
SP NCFS000 PP1CP00 VMIF1P0
DA0FS0 NCFS000 VMIF1P0
NCFS000 SP VMN0000 VMIS1P0

Esta transformación sí que puede suponer un importante cambio en los resultados del análisis estilométrico. Las obras cabe la posibilidad de que se relacionen por elementos léxicos concretos que pueden aparecer en varias obras por casualidad. De esta manera, si hay más de una obra palaciega, puede que palabras (y también sus correspondientes lemas) comunes sean *castillo*, *dama*, *princesa*, *conde*, *señor*, etc., mientras que otras comedias de corte mitológico puede que contengan palabras como *árboles*, *ninfa*, *río*, *montes*, *bosque*, etc. Esto puede hacer que las obras se relacionen por su temática más que por el factor que nos es de interés: su fecha de creación. Si estas palabras son transformadas a categorías gramaticales es posible que evitemos potencialmente este tipo de problemas y que florezcan otras relaciones diferentes.

Veamos cuáles son los resultados tras convertir el corpus en 270 etiquetas categoriales a través de *Freeling*. En este caso para el análisis se han empleado todas las etiquetas como MFW y unos parámetros 0% culling, palabras independientes, método Classic Delta. Los resultados pueden apreciarse en la Gráfica 4:

Datación automática teniendo en cuenta categorías gramaticales



Gráfica 4. Datación automática del corpus autógrafo de Lope teniendo en cuenta categorías gramaticales. 270 MFW, 0% culling, Classic Delta.

Los errores absolutos medios son de 1-NN = $\pm 5,4$; 2-NN = $\pm 4,7$; y 3-NN = $\pm 5,0$.

Los resultados, de nuevo, son bastante similares a los obtenidos con palabras y con lemas. Los tres elementos parecen comportarse de forma parecida cuando se trata de relacionar obras por su momento de escritura.

3.4 Comparación de resultados

En la Tabla 2 quedan reflejados los errores absolutos medios para cada uno de los análisis aquí acometidos.

	1-NN	2-NN	3-NN
Palabras	± 4,9 años	± 5,2 años	± 5,1 años
Lemas	± 4,9 años	± 5,2 años	± 5,2 años
Categorías gramaticales	± 5,4 años	± 4,7 años	± 5,0 años

Tabla 2. Comparación de los errores absolutos medios en las distintas pruebas.

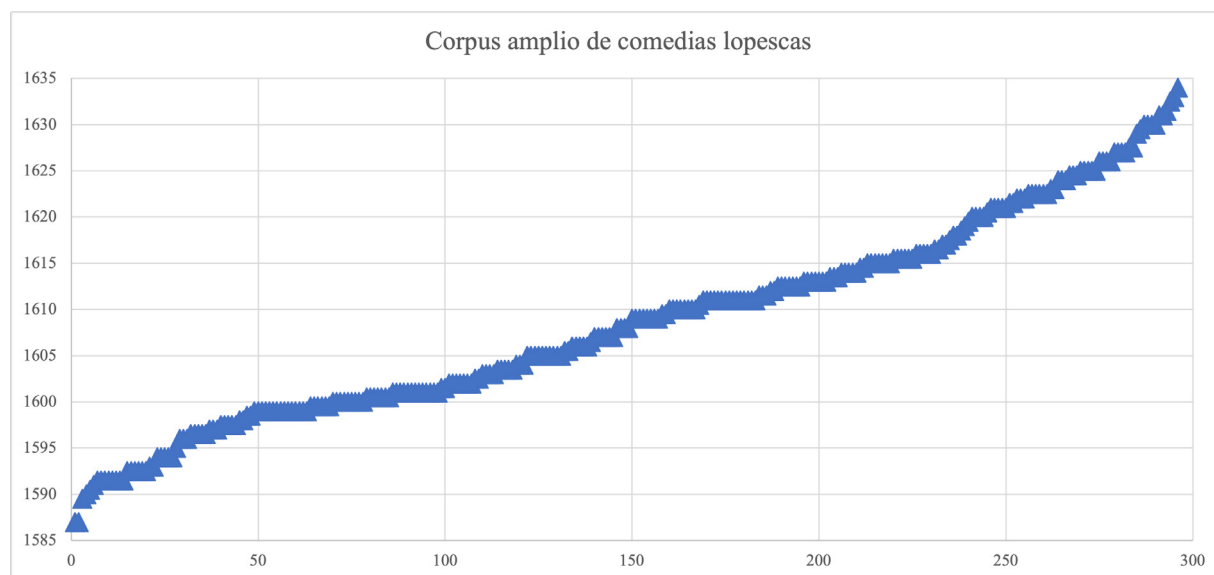
En los tres casos los resultados son muy similares: se rondan los ± 5 años de error. Las diferencias en un experimento tan reducido pueden, a mi juicio, despreciarse. Para el futuro queda explorar nuevas posibilidades con distintos elementos base o combinaciones de los aquí testados. También convendría probar sistemáticamente distintas configuraciones de parámetros (MFW, culling, etc.) en busca de las que mejores resultados ofrezcan para encontrar la huella temporal.

4. APLICACIÓN A UN CORPUS MÁS AMPLIO

Después del experimento con los autógrafos de datación segura de Lope de Vega, conviene ampliar varios órdenes la prueba para comprobar si sigue siendo efectiva. Para ello, tomo 296 obras de Lope de Vega de autoría segura (contrastando la información con Morley y Bruerton 1968, Vega García-Luengos 2021 y Cuéllar y Vega García-Luengos 2017-2022), con datación aproximada⁴ (siguiendo a Morley y Bruerton 1968) y de los que ha sido posible conseguir una edición digital adecuada. En los casos en los que contamos con una franja para la datación, tomo el valor central

4. Es preciso notar aquí que debido al elevado número de textos que manejamos para la prueba no ha podido acometerse una revisión pormenorizada sobre la datación tradicional de cada uno de los textos. Estas fechas han sido tomadas principalmente de Morley y Bruerton (aunque hay incongruencias en la datación de algunos textos en el propio monográfico entre la explicación de cada obra y las tablas de resumen finales) y completadas con varios estudios sugeridos en la base de datos Artelope.

como fecha para la obra. En la Gráfica 5 se puede apreciar la distribución de las comedias a lo largo de la producción lopesca.



Gráfica 5. Corpus amplio fechado de Lope de Vega.

Aunque en adelante tomaré estos años como fecha de escritura de las obras, están basados principalmente en el estudio de Morley y Bruerton, pero no contamos con el momento de escritura exacto de la mayoría de los textos. De hecho, como he apuntado previamente, aunque en ocasiones tengamos incluso la fecha de la pluma del propio Lope, no se puede descartar taxativamente que la obra hubiera sido escrita en otro año anterior. En este experimento se trata, por tanto, de partir de los datos lo más cercanos posible a la datación real, pero considerando siempre que los años aquí anotados son aproximaciones e intentos de datación, en la mayoría de casos provenientes del valor medio de un rango de tiempo ofrecido por los investigadores.

He aplicado en este caso las pruebas solo teniendo en cuenta palabras como elementos base. Como contamos con un mayor número de obras, utilizo para cada caso la primera obra más cercana (1-NN), la media de las dos obras más cercanas (2-NN), etc., así hasta tener en cuenta la media de las 10 obras más cercanas (10-NN). En la Tabla 3 podemos comprobar el error absoluto medio tras realizar estas

pruebas con las 296 comedias y los mismos parámetros explicados en los apartados anteriores (500 MFW, 0% culling, Classic Delta).

	Error absoluto medio
1-NN	± 5,2 años
2-NN	± 4,4 años
3-NN	± 4,1 años
4-NN	± 4,1 años
5-NN	± 4,0 años
6-NN	± 4,0 años
7-NN	± 4,0 años
8-NN	± 4,0 años
9-NN	± 3,9 años
10-NN	± 4,6 años

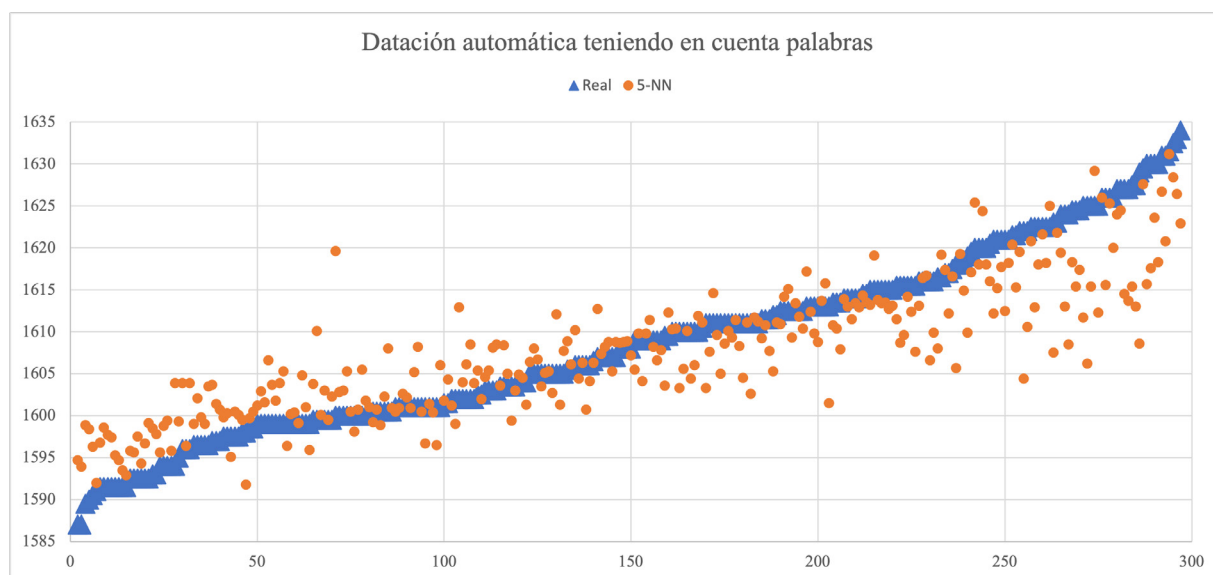
Tabla 3. Comparación de los errores absolutos medios al variar los vecinos más cercanos tenidos en cuenta.

Como podemos comprobar, los resultados han sido bastante acertados, consiguiendo un error absoluto medio en el entorno de ± 4 años,⁵ especialmente cuando utilizamos la media de entre las 5 y las 9 obras más cercanas en cada caso.

Podemos representar gráficamente los resultados de 9-NN, por ejemplo, para comprobar cómo se han comportado las obras en general. En la Gráfica 6 se aprecia que la tendencia en la datación automática de los textos se corresponde con la datación real de los mismos. Excepto algunas obras —como *La vida de San Pedro Nolasco*, la cual está fechada en 1629 pero nuestro proceso la data en 1610, o *Adonis y Venus*, la cual está fechada en 1600 mientras que nuestros análisis la datan en 1618,2—, el resto se comporta de forma bastante más ajustada. La mayoría, como se puede visualizar, se mueve en una franja inferior a ± 5 años, lo cual es un resultado muy valioso para nuestros propósitos. Como sucedía en los experimentos con autógrafos, las obras tempranas y las que corresponden a la mitad de la producción

5. En estos momentos estoy trabajando en mejorar estos resultados y, gracias a procesos de *machine learning*, en mi tesis *Propuestas de autoría y datación en el teatro de Lope de Vega a la luz de la Inteligencia Artificial* consigo datar el mismo corpus lopesco con errores en el entorno de ± 3 años.

lopesca se datan mejor que las de las últimas etapas. No es posible, como he explicado previamente, datar con seguridad las obras en los extremos absolutos, porque las medias nunca van a arrojar sus años como resultados.



Gráfica 6. Datación automática del corpus amplio de Lope teniendo en cuenta palabras y 9-NN. 500 MFW, 0% culling, Classic Delta.

Podemos ver en más detalle el experimento en la Tabla 4 (Datación tradicional, datación automática y error obtenido en un corpus amplio de obras lopescas. 500 MFW, 0% culling, Classic Delta),⁶ en la que ofrezco los resultados de las 296 obras utilizando la media de los nueve vecinos más cercanos (9-NN). En general, considero que los resultados son muy positivos, puesto que nos movemos en unos errores mínimos, incluso despreciables en los casos en los que partíamos ya de una horquilla de tiempo de la que hemos tomado el valor central para poder acometer los análisis. Observamos que los extremos son complicados de datar y que algunas obras, sobre todo las de género muy marcado, responden de forma inadecuada, porque tienden a relacionarse más por esta característica que por su fecha de composición. Es el caso de *Adonis* y *Venus*, respecto a la que ya Morley y Bruerton advertían sobre la dificul-

6. Para una explicación actualizada y más detallada sobre cada uno de los resultados de las obras aquí recogidas y de otras en la órbita lopesca, consultar los informes estilométricos en <https://etso.es/>.

tad que entrañaba su género literario para una correcta datación, o *La vida de San Pedro Nolasco*, *San Segundo* y *La niñez del padre Rojas*, de pronunciado género religioso y que responden mal a las pruebas. Es trabajo obligado el seguir avanzando en estos experimentos e intentar neutralizar las señales de género para quedarnos únicamente con la temporal. Cabe la posibilidad, también, de que en alguno de los casos en los que nuestra datación automática esté en disidencia con la datación tradicional, no se trate de un error y nos esté señalando con más acierto la datación real de la obra, extremo que habrá que comprobar con algunas comedias.

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Abanillo, El</i>	1615,0	1612,2	2,8
<i>Acero de Madrid, El</i>	1608,0	1604,8	3,2
<i>Adonis y Venus</i>	1600,0	1618,2	18,2
<i>Al pasar del arroyo</i>	1616,0	1611,7	4,3
<i>Alcaide de Madrid, El</i>	1599,0	1600,6	1,6
<i>Alcalde mayor, El</i>	1608,0	1607,6	0,4
<i>Almenas de Toro, Las</i>	1611,5	1610,2	1,3
<i>Amante agradecido, El</i>	1602,0	1601,7	0,3
<i>Amantes sin amor, Los</i>	1601,0	1600,2	0,8
<i>Amar como se ha de amar</i>	1622,5	1622,9	0,4
<i>Amar sin saber a quién</i>	1621,0	1615,3	5,7
<i>Amar, servir y esperar</i>	1629,5	1626,6	2,9
<i>Amigo hasta la muerte, El</i>	1611,0	1609,2	1,8
<i>Amigo por fuerza, El</i>	1599,0	1599,6	0,6
<i>Amistad y obligación</i>	1622,5	1617,8	4,7
<i>Amor con vista</i>	1626,0	1624,2	1,8
<i>Amor desatinado, El</i>	1597,0	1602,2	5,2
<i>Amor enamorado, El</i>	1630,0	1617,3	12,7
<i>Amor secreto hasta celos</i>	1614,0	1616,7	2,7

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Amor, pleito y desafío</i>	1621,0	1614,9	6,1
<i>Amores de Albanio y Ismenia, Los</i>	1592,5	1597,5	5,0
<i>Angélica en el Catay</i>	1599,0	1599,3	0,3
<i>Animal de Hungría, El</i>	1611,5	1608,8	2,7
<i>Anzuelo de Fenisa, El</i>	1605,0	1606,9	1,9
<i>Arauco domado, El</i>	1599,0	1603,3	4,3
<i>Arcadia, La</i>	1615,0	1617,9	2,9
<i>Arenal de Sevilla, El</i>	1603,0	1605,9	2,9
<i>Argel fingido y renegado de amor, El</i>	1599,0	1600,0	1,0
<i>Asalto de Matrique, por el príncipe de Parma, El</i>	1604,0	1603,8	0,2
<i>Ausente en el lugar, El</i>	1606,0	1606,8	0,8
<i>¡Ay, verdades que en amor...!</i>	1625,0	1614,1	10,9
<i>Bandos de Sena, Los</i>	1600,0	1603,2	3,2
<i>Barlaán y Josafat (Autógrafo)</i>	1611,0	1604,2	6,8
<i>Bastardo Mudarra, El</i>	1612,0	1603,9	8,1
<i>Batalla del honor, La</i>	1608,0	1608,6	0,6
<i>Batuecas del duque de Alba, Las</i>	1599,0	1607,1	8,1
<i>Bautismo del príncipe de Marruecos, El</i>	1599,0	1603,3	4,3
<i>Belardo el furioso</i>	1594,0	1596,4	2,4
<i>Bella Aurora, La</i>	1622,5	1613,2	9,3
<i>Bella malmaridada, La</i>	1596,0	1600,4	4,4
<i>Benavides, Los</i>	1600,0	1601,5	1,5
<i>Bizarrias de Belisa, Las</i>	1634,0	1621,4	12,6
<i>Blasón de los Chaves de Villalba, El</i>	1599,0	1600,6	1,6
<i>Boba para los otros y discreta para sí, La</i>	1630,0	1620,1	9,9

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Bobo del colegio, El</i>	1606,0	1611,0	5,0
<i>Boda entre dos maridos, La</i>	1599,5	1601,7	2,2
<i>Brasil restituido, El</i>	1625,0	1611,5	13,5
<i>Buena guarda, La</i>	1610,0	1608,1	1,9
<i>Burgalesa de Lerma, La</i>	1613,0	1614,2	1,2
<i>Burlas de amor, Las</i>	1591,0	1596,2	5,2
<i>Burlas veras, Las</i>	1624,5	1618,2	6,3
<i>Caballero de Illescas, El</i>	1602,0	1608,9	6,9
<i>Caballero de Olmedo, El</i>	1622,5	1618,4	4,1
<i>Caballero del milagro, El</i>	1593,0	1596,9	3,9
<i>Caballero del sacramento, El</i>	1610,0	1607,6	2,4
<i>Campana de Aragón, La</i>	1599,0	1603,0	4,0
<i>Capellán de la Virgen, El</i>	1615,0	1613,1	1,9
<i>Carbonera, La</i>	1624,5	1616,7	7,8
<i>Cardenal de Belén, El</i>	1610,0	1606,2	3,8
<i>Carlos Quinto en Francia</i>	1604,0	1602,6	1,4
<i>Casamiento en la muerte, El</i>	1595,0	1603,7	8,7
<i>Castelvines y Monteses</i>	1609,0	1604,9	4,1
<i>Castigo del discreto, El</i>	1599,5	1609,3	9,8
<i>Castigo sin venganza, El</i>	1631,0	1626,7	4,3
<i>Cautivos de Argel, Los</i>	1599,0	1604,3	5,3
<i>Celos de Rodamonte, Los</i>	1591,0	1596,9	5,9
<i>Cerco de Santa Fe, El</i>	1597,0	1602,3	5,3
<i>Comendadores de Córdoba, Los</i>	1597,0	1601,9	4,9
<i>Con su pan se lo coma</i>	1613,5	1609,6	3,9
<i>Conde Fernán González, El</i>	1611,0	1606,4	4,6

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Contienda de García de Paredes y el capitán Juan de Urbina, La</i>	1600,0	1604,1	4,1
<i>Contra valor no hay desdicha</i>	1627,5	1615,2	12,3
<i>Corona de Hungría y la injusta venganza, La</i>	1623,0	1610,2	12,8
<i>Corona merecida, La</i>	1603,0	1610,2	7,2
<i>Cortesía de España, La</i>	1610,0	1609,9	0,1
<i>Cuentas del Gran Capitán, Las</i>	1616,5	1607,9	8,6
<i>Cuerdo en su casa, El</i>	1607,0	1612,2	5,2
<i>Cuerdo loco, El</i>	1602,0	1602,3	0,3
<i>Dama boba, La</i>	1613,0	1611,9	1,1
<i>¿De cuándo acá nos vino?</i>	1613,0	1611,2	1,8
<i>Del mal lo menos</i>	1608,0	1609,4	1,4
<i>Del monte sale quien el monte quema</i>	1627,0	1621,1	5,9
<i>Desconfiado, El</i>	1615,5	1611,4	4,1
<i>Desdén vengado, El</i>	1617,0	1616,1	0,9
<i>Despertar a quien duerme, El</i>	1611,0	1606,6	4,4
<i>Desposorio encubierto, El</i>	1600,0	1601,7	1,7
<i>Desprecio agradecido, El</i>	1633,0	1623,7	9,3
<i>Devoción del rosario, La</i>	1605,0	1605,4	0,4
<i>Difunta pleiteada, La</i>	1594,0	1597,3	3,3
<i>Dios hace reyes</i>	1619,0	1610,7	8,3
<i>Discordia en los casados, La</i>	1611,0	1613,7	2,7
<i>Discreta enamorada, La</i>	1606,0	1606,6	0,6
<i>Divina vencedora, La</i>	1601,0	1603,2	2,2
<i>Divino africano, El</i>	1610,0	1607,0	3,0
<i>Dómine Lucas</i>	1592,5	1597,0	4,5

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Don Juan de Castro (Primera parte)</i>	1606,0	1605,9	0,1
<i>Don Juan de Castro (Segunda parte)</i>	1608,0	1604,4	3,6
<i>Don Lope de Cardona</i>	1611,0	1606,2	4,8
<i>Doncella Teodor, La</i>	1611,0	1606,3	4,7
<i>Duque de Viseo, El</i>	1608,5	1608,6	0,1
<i>Ello dirá</i>	1614,0	1613,3	0,7
<i>Embustes de Celauro, Los</i>	1600,0	1599,8	0,2
<i>Enemigo engañado, El</i>	1595,5	1596,8	1,3
<i>Enemigos en casa, Los</i>	1612,5	1610,9	1,6
<i>Envidia de la nobleza, La</i>	1614,0	1611,3	2,7
<i>Esclava de su galán, La</i>	1626,0	1616,8	9,2
<i>Esclavo de Roma, El</i>	1599,5	1600,6	1,1
<i>Esclavos libres, Los</i>	1599,0	1602,9	3,9
<i>Escolástica celosa, La</i>	1598,0	1597,3	0,7
<i>Espanoles en Flandes, Los</i>	1601,5	1603,6	2,1
<i>Estefanía la desdichada</i>	1604,0	1603,2	0,8
<i>Fábula de Perseo, La</i>	1611,0	1613,6	2,6
<i>Favor agradecido, El</i>	1593,0	1597,9	4,9
<i>Fe rompida, La</i>	1599,0	1601,4	2,4
<i>Felisarda, La</i>	1620,0	1615,6	4,4
<i>Ferias de Madrid, Las</i>	1587,0	1595,1	8,1
<i>Firmeza en la desdicha, La</i>	1611,0	1608,1	2,9
<i>Flores de don Juan y rico y pobre trocados, Las</i>	1613,5	1611,1	2,4
<i>Fortuna merecida, La</i>	1607,0	1609,4	2,4
<i>Francesilla, La</i>	1596,0	1602,4	6,4

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Fuenteovejuna</i>	1613,0	1607,5	5,5
<i>Fuerza lastimosa, La</i>	1602,0	1599,5	2,5
<i>Galán Castrucho, El</i>	1593,0	1594,4	1,4
<i>Galán de la Membrilla, El</i>	1615,0	1612,3	2,7
<i>Galán escarmentado, El</i>	1596,5	1599,2	2,7
<i>Gallarda toledana, La</i>	1602,0	1604,3	2,3
<i>Gallardo catalán, El</i>	1600,0	1604,0	4,0
<i>Ganso de oro, El</i>	1591,5	1595,3	3,8
<i>Genovés liberal, El</i>	1601,0	1608,6	7,6
<i>Gran duque de Moscovia y emperador perseguido, El</i>	1606,0	1602,6	3,4
<i>Grandezas de Alejandro, Las</i>	1606,0	1610,1	4,1
<i>Grao de Valencia, El</i>	1589,5	1600,6	11,1
<i>Guanches de Tenerife, Los</i>	1605,0	1603,6	1,4
<i>Guante de doña Blanca, El</i>	1632,5	1626,4	6,1
<i>Guerras de amor y de honor</i>	1615,5	1606,3	9,2
<i>Halcón de Federico, El</i>	1603,0	1608,3	5,3
<i>Hamete de Toledo, El</i>	1609,0	1607,7	1,3
<i>Hermosa Alfreda, La</i>	1599,0	1599,0	0,0
<i>Hermosa Ester, La</i>	1610,0	1607,6	2,4
<i>Hermosura aborrecida, La</i>	1607,0	1607,8	0,8
<i>Hidalgo Bencerraje, El</i>	1605,5	1604,7	0,8
<i>Hidalgos del aldea, Los</i>	1609,5	1612,3	2,8
<i>Hijo de los leones, El</i>	1621,0	1618,0	3,0
<i>Hijo de Reduán, El</i>	1591,5	1599,7	8,2
<i>Hijo sin padre, El</i>	1615,5	1611,9	3,6

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Hijo venturoso, El</i>	1591,5	1596,1	4,6
<i>Historia de Tobías, La</i>	1609,0	1606,9	2,1
<i>Hombre de bien, El</i>	1605,0	1605,9	0,9
<i>Hombre por su palabra, El</i>	1614,5	1614,5	0,0
<i>Honrado hermano, El</i>	1599,0	1602,9	3,9
<i>Imperial de Otón, La</i>	1597,0	1601,7	4,7
<i>Infanta desesperada, La</i>	1591,5	1595,8	4,3
<i>Ingratitud vengada, La</i>	1590,0	1596,5	6,5
<i>Ingrato arrepentido, El</i>	1600,0	1601,9	1,9
<i>Inocente Laura, La</i>	1606,0	1608,8	2,8
<i>Jorge Toledano</i>	1596,0	1595,9	0,1
<i>Juan de Dios y Antón Martín</i>	1607,0	1605,9	1,1
<i>Juez en su causa, El</i>	1610,0	1607,5	2,5
<i>Laberinto de Creta, El</i>	1613,5	1614,9	1,4
<i>Labrador venturoso, El</i>	1621,0	1615,2	5,8
<i>Laura perseguida</i>	1594,0	1602,5	8,5
<i>Leal criado, El</i>	1594,0	1597,5	3,5
<i>Limpieza no manchada, La</i>	1618,0	1605,0	13,0
<i>Llegar en ocasión, El</i>	1606,0	1604,7	1,3
<i>Lo cierto por lo dudoso</i>	1622,0	1619,7	2,3
<i>Lo fingido verdadero</i>	1608,0	1606,9	1,1
<i>Lo que está determinado</i>	1616,0	1613,7	2,3
<i>Lo que ha de ser</i>	1624,0	1616,6	7,4
<i>Lo que hay que fiar del mundo</i>	1610,0	1606,8	3,2
<i>Lo que pasa en una tarde</i>	1617,0	1612,2	4,8
<i>Loco por fuerza, El</i>	1602,5	1604,2	1,7

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Locos de Valencia, Los</i>	1592,5	1594,6	2,1
<i>Locos por el cielo, Los</i>	1600,5	1601,4	0,9
<i>Locura por la honra, La</i>	1611,0	1610,7	0,3
<i>Lucinda perseguida</i>	1600,5	1600,6	0,1
<i>Madre de la mejor, La</i>	1611,0	1605,6	5,4
<i>Maestro de danzar, El</i>	1594,0	1596,8	2,8
<i>Malcasada, La</i>	1612,5	1610,6	1,9
<i>Marido más firme, El</i>	1620,5	1616,8	3,7
<i>Mármol de Felisardo, El</i>	1596,0	1602,8	6,8
<i>Marqués de las Navas, El</i>	1624,0	1618,4	5,6
<i>Marqués de Mantua, El</i>	1596,0	1598,9	2,9
<i>Más galán portugués, Duque de Berganza, El</i>	1611,0	1611,3	0,3
<i>Más pueden celos que amor</i>	1627,0	1627,7	0,7
<i>Más valéis vos, Antona, que la corte toda</i>	1621,5	1621,9	0,4
<i>Mayor imposible, El</i>	1615,0	1613,5	1,5
<i>Mayor victoria, La</i>	1621,0	1619,0	2,0
<i>Mayor virtud de un rey, La</i>	1630,0	1623,2	6,8
<i>Mayorazgo dudoso, El</i>	1600,5	1599,7	0,8
<i>Mayordomo de la duquesa Amalfi, El</i>	1605,0	1606,8	1,8
<i>Mejor alcalde, el rey, El</i>	1621,5	1613,6	7,9
<i>Mejor maestro, el tiempo, El</i>	1611,0	1610,1	0,9
<i>Melindres de Belisa, Los</i>	1608,0	1611,1	3,1
<i>Mesón de la corte, El</i>	1591,5	1596,5	5,0
<i>Mirad a quién alabáis</i>	1620,0	1621,4	1,4
<i>Mudanzas de fortuna y sucesos de don Beltrán de Aragón, Las</i>	1607,5	1609,6	2,1

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Muertos vivos, Los</i>	1600,5	1598,5	2,0
<i>Mujeres sin hombres, Las</i>	1615,5	1611,0	4,5
<i>Nacimiento de Cristo, El (Parte XXIV)</i>	1614,0	1606,4	7,6
<i>Nacimiento de Ursón y Valentín, reyes de Francia, El</i>	1591,5	1593,8	2,3
<i>Nadie se conoce</i>	1618,0	1618,3	0,3
<i>Necedad del discreto, La</i>	1613,0	1612,9	0,1
<i>Negro del mejor amo, Antiobo de Cerdeña, El</i>	1601,0	1605,9	4,9
<i>Niña de plata, La</i>	1611,0	1608,9	2,1
<i>Niñez del padre Rojas, La</i>	1625,0	1612,6	12,4
<i>Niño inocente de La Guardia, El</i>	1603,0	1604,6	1,6
<i>No son todos rui señores</i>	1630,0	1619,8	10,2
<i>Noche de San Juan, La</i>	1631,0	1620,2	10,8
<i>Noche toledana, La</i>	1605,0	1605,2	0,2
<i>Mayor victoria de Alemania, La</i>	1622,0	1607,8	14,2
<i>Nuevo mundo descubierto por Cristóbal Colón, El</i>	1600,5	1604,1	3,6
<i>Obediencia laureada y primer Carlos de Hungría, La</i>	1605,5	1605,4	0,1
<i>Obras son amores</i>	1615,5	1613,9	1,6
<i>Ocasión perdida, La</i>	1601,0	1604,6	3,6
<i>Octava maravilla, La</i>	1609,0	1609,3	0,3
<i>Paces de los reyes y judía de Toledo, Las</i>	1611,0	1606,2	4,8
<i>Padrino desposado, El</i>	1599,0	1601,7	2,7
<i>Pastoral de Jacinto, La</i>	1597,5	1597,5	0,0
<i>Pedro Carbonero</i>	1603,0	1607,2	4,2
<i>Peligros de la ausencia, Los</i>	1616,5	1619,6	3,1

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Peribáñez y el comendador de Ocaña</i>	1606,5	1611,4	4,9
<i>Perro del hortelano, El</i>	1613,0	1613,7	0,7
<i>Perseguido, El</i>	1590,0	1600,2	10,2
<i>Piadoso aragonés, El</i>	1626,0	1620,2	5,8
<i>Piadoso veneciano, El</i>	1603,5	1601,1	2,4
<i>Pleitos de Inglaterra, Los</i>	1600,5	1605,8	5,3
<i>Pobrezas de Reinaldos, Las</i>	1599,0	1605,3	6,3
<i>Poder en el discreto, El</i>	1623,0	1621,1	1,9
<i>Poder vencido y amor premiado, El</i>	1614,0	1611,6	2,4
<i>Ponces de Barcelona, Los</i>	1611,0	1611,1	0,1
<i>Por la puente, Juana</i>	1624,5	1614,2	10,3
<i>Porceles de Murcia, Los</i>	1606,0	1603,7	2,3
<i>Porfiando vence amor</i>	1625,0	1627,1	2,1
<i>Porfiar hasta morir</i>	1626,0	1615,0	11,0
<i>Postrer godo de España, El</i>	1599,5	1603,6	4,1
<i>Prados de León, Los</i>	1605,0	1609,6	4,6
<i>Premio de la hermosura, El</i>	1612,5	1614,6	2,1
<i>Premio del bien hablar, El</i>	1624,5	1619,3	5,2
<i>Primer Fajardo, El</i>	1611,0	1603,4	7,6
<i>Primer rey de Castilla, El</i>	1600,5	1602,4	1,9
<i>Primera información, La</i>	1622,5	1620,0	2,5
<i>Príncipe despeñado, El</i>	1602,0	1602,2	0,2
<i>Príncipe inocente, El</i>	1590,0	1596,3	6,3
<i>Príncipe perfecto, El (Primera parte)</i>	1614,0	1613,7	0,3
<i>Príncipe perfecto, El (Segunda parte)</i>	1616,0	1614,6	1,4
<i>Prisión sin culpa, La</i>	1601,0	1600,0	1,0

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Prueba de los amigos, La</i>	1604,0	1601,9	2,1
<i>Prueba de los ingenios, La</i>	1612,5	1609,4	3,1
<i>Querer la propia desdicha</i>	1619,0	1617,3	1,7
<i>Quien más no puede</i>	1616,0	1610,4	5,6
<i>Quien todo lo quiere</i>	1620,0	1615,1	4,9
<i>Quinta de Florencia, La</i>	1600,0	1606,3	6,3
<i>Ramilletes de Madrid, Los</i>	1615,0	1612,3	2,7
<i>Remedio en la desdicha, El</i>	1596,0	1600,6	4,6
<i>Resistencia honrada y condesa Matilde, La</i>	1601,0	1604,8	3,8
<i>Robo de Dina, El</i>	1618,5	1615,5	3,0
<i>Roma abrasada</i>	1599,0	1599,9	0,9
<i>Saber puede dañar, El</i>	1622,5	1621,5	1,0
<i>San Segundo</i>	1594,0	1607,0	13,0
<i>Santa Liga, La</i>	1599,0	1602,4	3,4
<i>Santiago el verde</i>	1615,0	1612,4	2,6
<i>Secretario de sí mismo, El</i>	1605,0	1602,1	2,9
<i>Sembrar en buena tierra, El</i>	1616,0	1612,8	3,2
<i>Serrana de la Vera, La</i>	1596,5	1601,9	5,4
<i>Serrana de Tormes, La</i>	1592,5	1599,6	7,1
<i>Servir a buenos, El</i>	1622,5	1622,4	0,1
<i>Servir a señor discreto</i>	1611,0	1611,2	0,2
<i>Servir con mala estrella, El</i>	1605,0	1608,7	3,7
<i>Si no vieran las mujeres</i>	1631,5	1626,5	5,0
<i>Sin secreto no hay amor</i>	1626,0	1620,2	5,8
<i>Sol parado, El</i>	1599,5	1599,9	0,4
<i>Soldado amante, El</i>	1594,0	1600,1	6,1

Título	Datación tradicional	Datación automática (9-NN)	Error
<i>Sortija del olvido, La</i>	1612,5	1609,1	3,4
<i>Tellos de Meneses (Primera parte), Los</i>	1624,0	1610,4	13,6
<i>Testigo contra sí, El</i>	1605,5	1604,2	1,3
<i>Testimonio vengado, El</i>	1599,5	1601,1	1,6
<i>Tirano castigado, El</i>	1601,0	1600,4	0,6
<i>Torneos de Aragón, Los</i>	1597,0	1599,9	2,9
<i>Trabajos de Jacob, Los</i>	1625,0	1612,7	12,3
<i>Traición bien acertada, La</i>	1591,5	1594,0	2,5
<i>Tres diamantes, Los</i>	1601,0	1602,0	1,0
<i>Triunfo de la humildad y soberbia abatida, El</i>	1613,0	1604,1	8,9
<i>Valiente Céspedes, El</i>	1613,5	1608,4	5,1
<i>Valor de las mujeres, El</i>	1615,5	1611,7	3,8
<i>Valor, fortuna y lealtad de los Tellos de Meneses (Segunda parte)</i>	1627,5	1612,7	14,8
<i>Vaquero de Moraña, El</i>	1601,0	1601,4	0,4
<i>Varona castellana, La</i>	1599,0	1606,0	7,0
<i>Vellocino de oro, El</i>	1622,0	1614,7	7,3
<i>Vengadora de las mujeres, La</i>	1617,5	1619,0	1,5
<i>Venganza venturosa, La</i>	1611,5	1613,4	1,9
<i>Verdadero amante, El</i>	1587,0	1595,0	8,0
<i>Vida de san Pedro Nolasco, La</i>	1629,0	1610,6	18,4
<i>Villana de Getafe, La</i>	1612,0	1612,1	0,1
<i>Villano en su rincón, El</i>	1611,0	1611,7	0,7
<i>Viuda valenciana, La</i>	1599,5	1600,4	0,9
<i>Viuda, casada y doncella</i>	1597,0	1600,8	3,8

5. CONCLUSIONES

Después de los experimentos, primero sobre un corpus controlado de autógrafos y después sobre uno más amplio de comedias lopescas, parece plausible confirmar la hipótesis de partida: la estilometría es capaz de relacionar las obras teatrales de Lope de Vega por su momento de escritura. El uso de palabras, lemas y categorías gramaticales parece comportar resultados similares. Desde aquí invito a realizar trabajos que amplíen estas conclusiones, sigan arrojando luz a la datación en el teatro aurisecular e incorporen otros dramaturgos a las pruebas.

Los resultados son lo suficientemente acertados como para empezar, con precaución, a tenerlos en cuenta cuando enfrentemos un texto de Lope de Vega sin datar. Por supuesto, la fecha aquí obtenida no será la respuesta definitiva y es posible que falle por diversos motivos (como en casos de piezas con un género muy marcado), pero permitirá encuadrar la obra en unas franjas de tiempo sobre las que empezar a trabajar. La estilometría, al igual que sucede en el caso de la autoría, nos ofrece unas pistas y nos invita a explorarlas con todas las herramientas histórico-filológicas de las que dispongamos.

BIBLIOGRAFÍA

- CALVO TELLO, José, «Entendiendo Delta desde las humanidades», *Caracteres: estudios culturales y críticos de la esfera digital*, V 1 (2016), pp. 140-176.
- CALVO TELLO, José, *The Novel in the Spanish Silver Age. A Digital Analysis of Genre Using Machine Learning*, Bielefeld University Press, Bielefeld, 2021.
- CELMA VALERO, María Pilar, y Cristina RUIZ URBÓN, «Una aproximación a la escritura de Miguel Delibes desde la estilometría», *Tonos Digital*, 41 (2022).
- CEREZO SOLER, Juan, y José CALVO TELLO, «Autoría y estilo. Una atribución cervantina desde las humanidades digitales. El caso de *La conquista de Jerusalén*», *Anales Cervantinos*, LI (2019), pp. 231-250.
- CAMPIÓN LARUMBE, Miguel, y Álvaro CUÉLLAR, «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*, III (2021), pp. 59-69.
- CUÉLLAR, Álvaro, «Stylometry and Spanish Golden Age Theatre: An Evaluation of Authorship Attribution in a Control Group of One Hundred Undisputed Plays», en *Digital Stylistics in Romance Studies and Beyond*, eds. R. Hesselbach et al. (en prensa).
- CUÉLLAR, Álvaro, y Germán VEGA GARCÍA-LUENGOS, *ETSO: Estilometría aplicada al Teatro del Siglo de Oro, 2917-2022*, en línea, <<http://etso.es>>.
- DEMATTE, Claudia, «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*, XXXV 3 (2019), pp. 852-874.
- DEMATTE, Claudia, «El Segundo tomo de comedias de Juan Pérez de Montalbán: hacia un mapa estilométrico de las comedias auténticas, dudosas y ajenas», *Talía. Revista de estudios teatrales*, III (2021), pp. 71-78.
- EDER, Maciej, Jan RYBICKI y Mike KESTEMONT, «Stylometry with R: a package for computational text analysis», *R Journal*, VIII 1 (2016), pp. 107-121.
- GARCÍA-REIDY, Alejandro, «Deconstructing the Authorship of Siempre ayuda la verdad: A Play by Lope de Vega?», *Neophilologus*, CIII 4 (2019), pp. 493-510.
- MADROÑAL, Abraham, ed., Lope de Vega, *El renacer del Fénix: Yo he hecho lo que he podido, Fortuna lo que ha querido*, Universidad de Valladolid, Valladolid, 2021.
- MARCOS RODRÍGUEZ, Emma María, «Texto, atribución y censura de *Próspera y Adver-*

- sa fortuna de Don Alvaro de Luna», *Talía. Revista de estudios teatrales*, III (2021), pp. 79-89.
- MARTÍNEZ CARRO, Elena, «Límites estilométricos en una miscelánea áurea: *Favores de las musas* de Sebastián Francisco de Medrano», *Hipogrifo*, IX 1 (2021), pp. 159-174.
- MARTÍNEZ CARRO, ELENA, «*David perseguido y montes de Gelboé*, novedades estilométricas ante una atribución dudosa», *Anuario Lope de Vega. Texto, literatura, cultura*, XXVIII (2022), pp. 401-422.
- PADRÓ, Lluís, y Evgeny STANILOVSKY, «FreeLing 3.0: Towards Wider Multilinguality», en *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, 2012.
- PRESOTTO, Marco, *Le commedie autografe di Lope de Vega: catalogo e studio*, Reichenberger, Kassel, 2000.
- RODRÍGUEZ-GALLEGO, Fernando, ed., Agustín Moreto, *La adúltera penitente*, Biblioteca Virtual Miguel de Cervantes, Alicante, 2019.
- MORLEY, Sylvanus Griswold, y Courtney BRUERTON, *Cronología de las comedias de Lope de Vega. Con un examen de las atribuciones dudosas, basado todo ello en un estudio de versificación estrófica*, Gredos, Madrid, 1968.
- PEÑARRUBIA NAVARRO, Pablo, «Estilometría con fines geolingüísticos aplicada al corpus COSER», *Revista de Humanidades Digitales*, VI (2021), pp. 22-42.
- O'SULLIVAN, James, «The sociology of style: Writing and influence within literary families», *Poetics* XCII A (2022), pp. 1-6.
- ULLA LORENZO, Alejandra, Elena MARTÍNEZ CARRO y José CALVO TELLO, «Las comedias de dudosa atribución de Agustín Moreto: nuevas perspectivas estilométricas», *Neophilologus*, CV 1 (2021), pp. 57-73.
- VEGA GARCÍA-LUENGOS, Germán, «Juan Ruiz de Alarcón recupera *La monja alférez*», en *Sor Juana Inés de la Cruz y el teatro novohispano. XLII Jornadas de teatro clásico de Almagro*, eds. R. González Cañal y A. García González, Ediciones de la Universidad de Castilla-La Mancha, [Cuenca], 2021, pp. 1-48, 2020.
- VEGA GARCÍA-LUENGOS, Germán, «Las comedias de Lope de Vega: confirmaciones de autoría y nuevas atribuciones desde la estilometría (I)», *Talía. Revista de estudios teatrales*, III (2021), pp. 91-108.