

La estilometría rítmica a examen en un corpus de comedias seguras de Lope de Vega*

Testing Rhythmic Stylometry with a Corpus of Lope de Vega's Securely Attributed Comedias

ÁLVARO CUÉLLAR

Universitat Autònoma de Barcelona, Edifici A - Campus Universitari UAB, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, España.

Dirección de correo electrónico: alvaro.cuellar@uab.cat.

ORCID: <https://orcid.org/0000-0002-9934-6321>.

Recibido/Received: 16-1-2026. Aceptado/Accepted: 4-4-2026.

Cómo citar/How to cite: Cuéllar, Álvaro (2026). "La estilometría rítmica a examen en un corpus de comedias seguras de Lope de Vega". *Castilla. Estudios de Literatura*, 17, pp. 114-135. DOI: <https://doi.org/10.24197/6dkssk45>.

Artículo de acceso abierto distribuido bajo una [Licencia Creative Commons Atribución 4.0 Internacional \(CC-BY 4.0\)](#). / Open access article under a [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](#).

Resumen: Este artículo evalúa la capacidad discriminativa de la estilometría rítmica en un entorno controlado para la atribución de autoría en el teatro del Siglo de Oro. A partir de un corpus binario de comedias seguras de Lope de Vega frente a un conjunto amplio de comedias no lopescas, las obras se representan mediante frecuencias de patrones acentuales obtenidos por escansión automática. Se aplican modelos supervisados de clasificación y se observa que el grado de acierto se concentra en una ventana de patrones frecuentes, aproximadamente entre 100 y 200, alcanzando valores de F1 cercanos al 95%. Aunque los resultados son prometedores, el análisis de errores revela falsos negativos recurrentes, incluidos dos autógrafos lopescos, lo que señala margen de mejora y aconseja prudencia interpretativa al aplicar el método a casos disputados.

Palabras clave: estilometría rítmica; atribución de autoría; Lope de Vega; teatro del Siglo de Oro; aprendizaje automático

Abstract: This article evaluates the discriminative capacity of rhythmic stylometry in a controlled setting for authorship attribution in Spanish Golden Age theatre. Using a binary corpus of securely

* Este trabajo se ha realizado en el marco de los proyectos de investigación ISTAE: Impresos sueltos del teatro antiguo español (II) (PID2022-136431NB-C66); La integral dramática de Lope de Vega (II) (PID2024-155554NBI00); METADRAMA: Teatro áureo en diacronía: estudio, edición y puesta en escena del verso clásico (PID2024-161619NA-I00); Thal-IA: Patrimonio teatral áureo: Inteligencia Artificial y Fotografía Espectral (CNS2023-145014); y Tracing Regularities in Pedro Calderón de la Barca's Dramatic Oeuvre with a Computational Approach (DFG Project No. 508056339).

attributed comedias by Lope de Vega contrasted with a large set of non-Lope plays, works are represented through frequencies of stress-based rhythmic patterns obtained via automatic scansion. Supervised classification models are applied, and performance is shown to concentrate in a window of frequent patterns, approximately between 100 and 200, reaching F1 values close to 95%. Although the results are promising, error analysis reveals recurrent false negatives, including two autograph plays, indicating room for improvement and recommending interpretive caution when applying the method to disputed cases.

Keywords: rhythmic stylometry; authorship attribution; Lope de Vega; Spanish Golden Age theatre; machine learning

INTRODUCCIÓN

La estilometría rítmica se está consolidando en los últimos años como una vía metodológica alternativa y complementaria para los estudios de autoría en el teatro del Siglo de Oro, al centrar el análisis en niveles formales vinculados a la prosodia y a la organización acentual del verso, más allá del contenido léxico de los textos (Kroll, 2022; Kroll y Sanz-Lázaro, 2022; Kroll y Sanz-Lázaro, 2023; Sanz-Lázaro, 2022; Sanz-Lázaro, 2023a; Sanz-Lázaro, 2023b). La atribución de la autoría continúa siendo un problema central en el campo, tanto por la magnitud del repertorio conservado como por la frecuencia de atribuciones dudosas, colaboraciones y procesos de transmisión textual complejos. En este contexto, la estilometría léxica ha demostrado una eficacia muy elevada y constituye hoy una herramienta fundamental, pero presenta limitaciones bien identificadas: su sensibilidad a la temática y al género, la interferencia de la variación diacrónica del estilo, la dificultad para discriminar autores con perfiles léxicos próximos, los problemas derivados de la autoría compartida y la posible distorsión introducida por la tradición editorial y ortográfica de los textos. La estilometría rítmica, al trabajar con la secuencia y la frecuencia de patrones acentuales, aborda el problema desde una dimensión formal distinta y puede resultar menos dependiente de algunos de estos factores, al menos en determinados contextos analíticos. Desde esta perspectiva, el estudio sistemático del ritmo no pretende sustituir la estilometría léxica, sino aportar una señal independiente que permita contrastar, matizar o reforzar hipótesis de autoría en casos concretos y metodológicamente acotados.

En un reciente trabajo (Cuéllar, 2025a) se articuló este desplazamiento metodológico desde la estilometría léxica —centrada en frecuencias de palabras— hacia una estilometría rítmica basada en la escansión

automática y en la codificación computacional de la alternancia de sílabas tónicas y átonas, interpretada como una onda prosódica que recorre el texto dramático (Tabla 1).

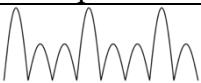
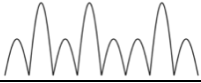
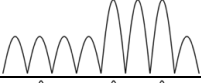
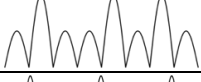
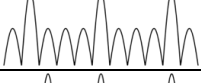
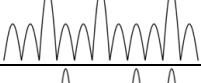
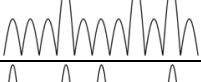
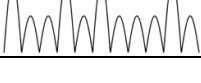
Verso	Onda prosódica	Patrón rítmico
dame herreruero y sombrero		+--+--+
con esta media sotana		-+-+--+
aunque mi ignorancia es llana		----+++
cogerte a las manos quiero		-+--++-
ahora que descansan los aceros		-+----+--
de mis nunca olvidados compañeros		--+--+--+
y el enemigo de furor armado		---+--+--
huye vencido o duerme fatigado		+--+--+--

Tabla 1. Ejemplos de octosílabos y endecasílabos transformados en ondas prosódicas y sus correspondientes patrones rítmicos (Cuéllar, 2025a)

El estudio se planteó como un experimento controlado con un corpus de cien comedias de autoría segura atribuidas a nueve dramaturgos, con el fin de evaluar la capacidad discriminativa de los patrones rítmicos extraídos automáticamente y compararla con la de los enfoques léxicos habituales. A partir de la extracción masiva de patrones y su uso como rasgos estilométricos, se aplicaron procedimientos de clasificación y validación cruzada mediante técnicas de aprendizaje automático. Los resultados mostraron que, cuando la tarea se formula como clasificación multiclase, el rendimiento rítmico es sistemáticamente inferior al léxico, que alcanza niveles cercanos al techo en este tipo de corpus una vez se supera un umbral de palabras frecuentes. Esta diferencia no se interpretó como una refutación del enfoque rítmico, sino como un indicio de su

mayor sensibilidad al ruido de la escansión automática, a la polimetría y a la heterogeneidad formal de las obras, factores que afectan de manera desigual a la estabilidad de la señal. Un resultado especialmente relevante fue la reformulación del problema en términos binarios. En lugar de decidir entre varios autores simultáneamente, el experimento mostró que la oposición de un autor frente al resto incrementa de forma sustantiva la fiabilidad del sistema; en particular, la distinción obras lopescas versus obras no lopescas llegó a alcanzar un 100% de acierto bajo condiciones controladas, y los mejores comportamientos tendieron a concentrarse en un conjunto relativamente acotado de parámetros (en el entorno de 100 patrones rítmicos más frecuentes). Este resultado sugiere que la señal autoral rítmica, cuando existe, no se fortalece necesariamente por acumulación de rasgos raros, sino por la explotación de un núcleo de hábitos prosódicos altamente recurrentes. Ahora bien, el propio estudio también mostró que el ritmo no ofrece siempre una señal igualmente estable, incluso en obras canónicas y de autoría indiscutida. Un ejemplo ilustrativo es *La dama boba*, que en la red estilométrica rítmica se separa del conjunto principal de Lope y aparece en una posición distante, una divergencia que se ha relacionado de manera tentativa con su elevada densidad de cambios métricos y, por tanto, con efectos de polimetría sobre la representación agregada de patrones (Antonucci, 2020).

Este tipo de comportamiento obliga a extremar la cautela interpretativa: mientras que la estilometría léxica suele mantener una robustez alta en múltiples autores y escenarios, el rendimiento rítmico depende más estrechamente de condiciones formales y de calidad de anotación, y puede producir salidas menos consistentes si no se controla adecuadamente el ruido o la variación métrica. Sobre esta base, el objetivo del presente artículo es dar un paso adicional en la estabilización y el control de la señal rítmica. Para ello se construye un corpus binario más amplio —con un subconjunto de obras seguras de Lope y un conjunto de contraste de comedias no lopescas— y se aplica una batería de pruebas destinada a observar (i) hasta qué punto la estilometría rítmica discrimina de manera fiable en condiciones de control y (ii) qué características presentan los casos que se apartan del comportamiento esperado.

1. PREPARACIÓN DEL CORPUS DE CONTROL

Con el fin de evaluar la eficacia de la estilometría rítmica en un entorno empírico exigente —esto es, con un corpus amplio, heterogéneo y

representativo del teatro áureo—, el estudio se centra en Lope de Vega. La elección no responde únicamente a su centralidad canónica, sino a un criterio metodológico: la existencia de un repertorio particularmente extenso y bien estudiado, que facilita la construcción de un conjunto de control con suficiente variación interna y, al mismo tiempo, con un número elevado de obras comparables frente a un conjunto amplio de textos no lopescos. Este diseño resulta especialmente adecuado para una formulación binaria del problema (Lope / no Lope), que, como se ha señalado en trabajos previos, tiende a maximizar la detectabilidad de la señal rítmica frente a escenarios multiclase.

La preparación de un corpus rítmico requiere, sin embargo, un paso previo costoso: la escansión de los versos y su conversión a representaciones formales de acento y ritmo. Aunque en los últimos años se han desarrollado herramientas automáticas (Marco et al., 2021; Marco y Gonzalo, 2021; Navarro-Colorado, 2018; Sanz-Lázaro, 2024) capaces de realizar este proceso a gran escala —con aproximaciones complementarias y diferentes supuestos lingüísticos y métricos—, la escansión sigue siendo un punto crítico por su sensibilidad a irregularidades textuales, polimetría, versos partidos, deturpaciones de la tradición impresa y fenómenos ortológicos que afectan al cómputo silábico y a la asignación de acentos. En consecuencia, la creación *ex novo* de un corpus rítmico suficientemente amplio para experimentación controlada implica no solo la ejecución automática del escandido, sino también tareas de inspección y corrección que resultan difíciles de sostener cuando el objetivo es trabajar con multitud de obras.

Por este motivo, el experimento parte de materiales ya preparados en el marco del proyecto *Sound and Meaning*, en particular de la base de datos *Versos y estructuras teatrales áureos: A corpus of Spanish plays with dramatic and metric information*, que ofrece un recurso estructurado y reutilizable para análisis cuantitativos del verso (Kroll y Sanz-Lázaro, 2024). La base reúne versos procedentes de 425 obras escritas por o atribuidas a 23 dramaturgos, con una representación especialmente abundante de Lope de Vega y Calderón de la Barca, y utiliza como fuente textos tomados de ediciones filológicas acreditadas, que se consignan explícitamente en los campos de referencia editorial. Cada verso incorpora, además, una capa rica de metadatos: información editorial (autor, título, género/subgénero, identificación de la edición, etc.), información dramática (acto, parlamento e identificación del personaje o personajes que lo pronuncian, junto con su género masculino o femenino)

e información métrica extraída automáticamente (número de sílabas, núcleos silábicos, asonancia, consonancia y datos de ritmo). En el contexto del presente trabajo, esta última dimensión es crucial, ya que permite operar directamente con patrones rítmicos a escala de corpus y construir, con un coste de preparación significativamente menor, un conjunto de control amplio para contrastar el perfil rítmico de Lope frente a un repertorio extenso de dramaturgia coetánea no lopesca. Conviene añadir, no obstante, que la base presenta algunos inconvenientes menores para su explotación directa —por ejemplo, la presencia puntual de registros duplicados y ciertas discontinuidades en el enlace entre metadatos y versos—, de modo que requiere una depuración mínima previa; una vez realizado ese ajuste, el recurso ofrece datos de gran valor para el tipo de análisis cuantitativo aquí planteado.

A este corpus de referencia se incorporaron las obras depositadas en Zenodo (Cuéllar, 2025b) que sirvieron al experimento anteriormente mencionado, con el objetivo de aumentar el número total de textos y mejorar la cobertura comparativa, evitando duplicados cuando existía solapamiento con el corpus inicial. Para mantener un entorno de prueba lo más controlado posible, se depuró el conjunto eliminando aquellas piezas que no pertenecen al género de la comedia, excluyendo autos sacramentales, entremeses, loas y otras formas breves o piezas menores que, por su estatuto genérico y su distinta configuración formal, pueden introducir variación no deseada en la señal rítmica. En paralelo, y con el fin de maximizar la seguridad de la etiqueta autoral en el conjunto de control, se revisó la atribución de todas las obras y se restringió la selección a aquellas consideradas seguras en la clasificación de referencia de Morley y Bruerton (1968), apoyándose también operativamente en la información consignada en la base de datos Artelope.

Esta decisión metodológica tiene consecuencias directas sobre algunos casos en los que la evidencia cuantitativa reciente no coincide con la tradición crítica. Así, por ejemplo, *La reina Juana de Nápoles* se mantuvo en el corpus por figurar como obra segura en Morley y Bruerton, pese a que determinados análisis estilométricos de base léxica la sitúan fuera del repertorio lopesco o, cuando menos, problematizan su adscripción (Vega García-Luengos, 2021; Cuéllar, 2022). La inclusión de este tipo de piezas responde a un criterio de control experimental: si el objetivo es evaluar la eficacia de la estilometría rítmica como señal independiente, excluir de antemano —por motivos derivados de la estilometría léxica— obras que la crítica ha considerado seguras introduce

un sesgo circular y entrena el experimento para reproducir, por construcción, el veredicto del léxico. Mantener estos casos permite, en cambio, medir con mayor limpieza la estabilidad real de la señal rítmica y, cuando surjan discrepancias, convertirlas en un punto de análisis (polimetría, transmisión textual, heterogeneidad formal) en lugar de neutralizarlas mediante filtrado previo. El resultado de este proceso de integración y depuración es un subcorpus final de 110 comedias de Lope de Vega con autoría segura, que constituye el bloque lopesco sobre el que se articula la evaluación binaria del presente estudio. En lo que respecta al conjunto no lopesco, el corpus reúne 216 comedias correspondientes a 20 dramaturgos distintos. Este bloque de contraste cumple una función estrictamente experimental: proporcionar al clasificador una base heterogénea sobre la que aprender, por oposición, el perfil rítmico característico de Lope de Vega frente a un repertorio amplio de escritura coetánea. Conviene, no obstante, tener en cuenta que el conjunto presenta una sobrerrepresentación de Calderón de la Barca —derivada de las prioridades del proyecto *Sound and Meaning*—; pese a ello, se optó por no intervenir en esa distribución y mantenerla en el diseño experimental, por no prescindir de ninguna comedia que pudiera enriquecer el modelo.

Con el objetivo de facilitar la replicación y la mejora futura del trabajo, el conjunto de escansiones empleado en este estudio, integrado a partir de materiales procedentes de recursos previos (Kroll y Sanz-Lázaro, 2024; Cuéllar, 2025b), se encuentra disponible en Zenodo (Cuéllar, 2026; DOI: <https://doi.org/10.5281/zenodo.19418278>).

A continuación, se presentan (Tabla 2 y Tabla 3) los 50 patrones rítmicos más frecuentes calculados, por separado, para el subcorpus lopesco y para el conjunto de otros autores. Estas frecuencias ofrecen una primera aproximación descriptiva a la distribución de ritmos que maneja el experimento y permiten visualizar, incluso antes del modelado, la concentración de la señal en un núcleo reducido de patrones altamente recurrentes, frente a una larga cola de formas raras más expuestas a ruido y variación. Aunque este tipo de tablas no siempre es inmediatamente interpretable desde una lectura filológica tradicional, constituye una representación informativa esencial para el aprendizaje automático, ya que los modelos operan precisamente sobre diferencias sistemáticas —por pequeñas que sean— en la distribución relativa de estos rasgos entre conjuntos autorales.

Nº	Ritmo	Porcentaje
1º	-+-+--+	6.09%
2º	-+--++-	4.93%
3º	--+---+	4.63%
4º	---+--+	4.56%
5º	+--+--+	4.17%
6º	--+--+	3.26%
7º	-+-+--+	3.20%
8º	+--+---	2.80%
9º	-+-----	2.39%
10º	+--+--+	2.37%
11º	-+--++	2.26%
12º	+--+--+	2.19%
13º	--+---+	2.09%
14º	---+--+	2.02%
15º	+--+--+	1.87%
16º	++--++	1.70%
17º	----+--+	1.57%
18º	--+--+	1.46%
19º	+--+---	1.42%
20º	++--++	1.27%
21º	-+-----	1.25%
22º	--++--+	1.20%
23º	-++--+	1.19%
24º	-++--+	1.10%
25º	+--+--+	1.08%
26º	-+----+	1.03%
27º	++--++	0.96%
28º	+--+--+	0.94%
29º	-+-+---+--	0.94%
30º	-+---+---+	0.92%
31º	+---+--	0.91%
32º	--+--+---+	0.83%
33º	+--+--+	0.78%
34º	-+-+---+	0.77%
35º	-+-+--+--+	0.67%
36º	--++--+	0.67%
37º	---+--+	0.66%
38º	++--++	0.65%
39º	----+--	0.62%
40º	++-----	0.61%
41º	---+---+--	0.60%
42º	-++---+	0.59%
43º	-+---+--+	0.59%
44º	+--+--+	0.58%
45º	-+++--+	0.58%
46º	-+-+--+	0.56%
47º	+--+---+--+	0.55%
48º	-++--+	0.53%

49º	---+-+---+-	0.52%
50º	-+-++-+	0.51%

Tabla 2. 50 patrones rítmicos más frecuentes en el subcorpus lopesco, con su porcentaje de aparición sobre el total de versos analizados en ese conjunto.

Nº	Ritmo	Porcentaje
1º	--+---+-	6.61%
2º	-+-+--+-	6.46%
3º	-+-+--+	5.64%
4º	---+--+-	5.53%
5º	+--+--+-	4.24%
6º	--+-+--+	4.00%
7º	-+----+-	3.33%
8º	+--+---+-	3.23%
9º	+--+--+	2.64%
10º	-+-+--+	2.39%
11º	--+---+	2.12%
12º	----+-+	1.95%
13º	+---+--+	1.92%
14º	-+-+--+	1.82%
15º	---+--+	1.73%
16º	+--+--+	1.57%
17º	--++--+	1.56%
18º	-++---+-	1.49%
19º	++-+--+	1.43%
20º	--+-+--+	1.30%
21º	-++-+--+	1.25%
22º	-+----+	1.22%
23º	+--+---+	1.18%
24º	-+-++-+-	1.18%
25º	++--+--+	1.18%
26º	+--+--+	1.01%
27º	+--+--+	0.90%
28º	---++-+-	0.86%
29º	+--+--+	0.80%
30º	--+---+	0.74%
31º	-+-+--+	0.74%
32º	+---+--+	0.67%
33º	++----+-	0.65%
34º	--++--+	0.63%
35º	++-+--+	0.61%
36º	-+++--+	0.59%
37º	---+--+	0.58%
38º	+--+--+	0.58%
39º	----++	0.57%
40º	-++---+	0.56%
41º	--+---+-	0.56%

42°	-+---+---+-	0.49%
43°	+--+---+-	0.48%
44°	++--+--+	0.47%
45°	+-----+-	0.47%
46°	-+---+++-	0.46%
47°	-++-+-+	0.45%
48°	-+-++-+	0.45%
49°	+---+-+	0.42%
50°	-+---++-	0.41%

Tabla 3. 50 patrones rítmicos más frecuentes en el subcorpus no lopesco, con su porcentaje de aparición sobre el total de versos analizados en ese conjunto.

2. PRUEBAS Y EVALUACIÓN DE LOS RESULTADOS

Para someter la estilometría rítmica a un examen sistemático, el estudio se formula como un problema de clasificación supervisada mediante modelos clásicos de *machine learning*. En este enfoque, los algoritmos aprenden a distinguir entre dos grupos predefinidos —en nuestro caso, Lope frente a no Lope— a partir de un conjunto de rasgos cuantificables. Aquí, esos rasgos están constituidos por la distribución de los patrones rítmicos en cada obra, es decir, por la frecuencia relativa con la que el texto emplea los distintos esquemas de acentuación seleccionados.

La construcción de esos rasgos parte, primero, de identificar cuáles son los patrones rítmicos más frecuentes del corpus y fijar un umbral que determina cuántos de ellos se incorporan al análisis. A continuación, cada obra se representa como un perfil numérico que recoge la frecuencia con la que aparecen esos patrones. Para que estas frecuencias sean comparables entre obras y no queden distorsionadas por diferencias globales —por ejemplo, por la longitud del texto o por variaciones generales en el repertorio métrico—, se aplica una estandarización mediante *z-scores*, un procedimiento habitual en tareas de clasificación: en lugar de trabajar con la frecuencia en bruto, se expresa cada patrón en términos de cuánto se desvía su uso en una obra respecto de su comportamiento medio en el conjunto de referencia. En el esquema de validación *leave-one-out* empleado, esta estandarización se recalcula en cada iteración exclusivamente a partir del conjunto de entrenamiento, y los parámetros obtenidos (media y desviación típica de cada rasgo) se aplican después a la obra reservada para prueba. Los *z-scores* permiten poner todos los rasgos en una escala común y reducen el riesgo de que algunos patrones

pesen más que otros solo por razones de escala, contribuyendo así a estabilizar la señal estilométrica. Una vez realizada esta transformación, cada obra queda descrita por un vector de características cuya dimensión depende del número de patrones incluidos. El experimento consiste en aplicar una batería de modelos supervisados de clasificación para comparar su rendimiento y determinar (i) qué enfoque discrimina mejor la oposición Lope/no Lope y (ii) en qué rango de patrones rítmicos frecuentes se concentra la señal más informativa, antes de que la incorporación de patrones raros introduzca ruido. Para ello se ensayan cinco métodos clásicos y ampliamente utilizados en tareas de clasificación: SVM con kernel RBF, SVM lineal, regresión logística, Naive Bayes y k-Nearest Neighbors (k-NN). Esta selección es metodológicamente pertinente porque cubre supuestos y mecanismos de decisión complementarios. En particular, SVM lineal y regresión logística son clasificadores lineales: aprenden una función discriminante lineal (un hiperplano) en el espacio de rasgos. SVM con kernel RBF introduce no linealidad mediante una transformación implícita del espacio (kernel), lo que permite fronteras de decisión más flexibles. Naive Bayes constituye un enfoque probabilístico generativo: modela la distribución de los rasgos condicionada a la clase bajo un supuesto de independencia condicional. Por último, k-NN es un método no paramétrico basado en vecindad, que asigna la clase a partir de la proximidad a ejemplos de entrenamiento en el espacio de rasgos. En conjunto, esta diversidad permite evaluar si la señal rítmica queda bien capturada por separaciones lineales o si se beneficia de relaciones no lineales, y contrastarla con aproximaciones generativas y basadas en distancia. Todo el flujo del experimento (preprocesado, modelos, validación *leave-one-out* y métrica F1) puede implementarse de forma directa en Python con la librería *scikit-learn* (Pedregosa et al., 2011), que ya incluye los clasificadores y utilidades necesarias. Salvo indicación contraria, los modelos se han implementado con los valores de hiperparámetros por defecto de la librería.

La evaluación del rendimiento se realiza mediante validación cruzada *leave-one-out*: en cada iteración se extrae una obra, se entrena el modelo con todas las demás y se intenta clasificar la obra retirada; el proceso se repite hasta que cada texto ha actuado una vez como prueba. Los resultados se resumen con la medida F1, muy utilizada en clasificación porque equilibra dos dimensiones complementarias: por un lado, cuántas de las obras clasificadas como Lope lo son realmente (precisión) y, por otro, cuántas de las obras que son Lope se detectan efectivamente como tales

(exhaustividad). En este trabajo se toma la clase Lope como positiva, por lo que la medida F1 se calcula con respecto a su identificación frente al conjunto no lopesco, combinando precisión y exhaustividad sobre dicha clase. Este equilibrio es pertinente en un problema binario de atribución, donde interesa valorar simultáneamente el coste de aceptar falsos positivos y el de rechazar falsos negativos. Finalmente, aunque en otros diseños se separa un conjunto externo de prueba, aquí el esquema *leave-one-out* cumple esa función de forma interna: cada obra se evalúa siempre fuera del entrenamiento de su iteración y, dado que el corpus no se altera tras su selección, se evita que el experimento quede condicionado por ajustes retrospectivos orientados a optimizar el resultado.

Procedemos, por tanto, a ejecutar la batería experimental con cinco clasificadores distintos, registrando su comportamiento a medida que se varía de forma sistemática el número de patrones rítmicos más frecuentes (MFR) incluidos en la representación de cada obra. Los resultados se reportan mediante la medida F1 en su escala estándar de 0 a 1, donde 1 corresponde a una clasificación perfecta y 0 a un desempeño nulo. Esta presentación permite observar con claridad tanto qué clasificadores se comportan de manera más estable como en qué rangos de MFR la señal rítmica resulta más informativa. La Figura 1 muestra, de forma desagregada, las curvas de F1 para cada uno de los cinco métodos en función del umbral de MFR, lo que permite comparar directamente su estabilidad y sus máximos locales. Complementariamente, la Figura 2 sintetiza el comportamiento conjunto de la batería: representa la media del F1 entre clasificadores y una banda que recoge su dispersión (rango min-max) en cada valor de MFR, ofreciendo una lectura compacta del nivel de acuerdo entre modelos y de la existencia —o no— de una ventana de patrones especialmente favorable.

Como puede observarse, la banda de resultados es globalmente positiva: en una proporción amplia de configuraciones los modelos alcanzan valores de F1 superiores a 0,9 y, en los mejores casos, se aproximan a 0,95, lo que sugiere que, formulada como tarea binaria (Lope/no Lope), la señal rítmica contiene información discriminativa suficiente para sostener clasificaciones fiables en un entorno de control.

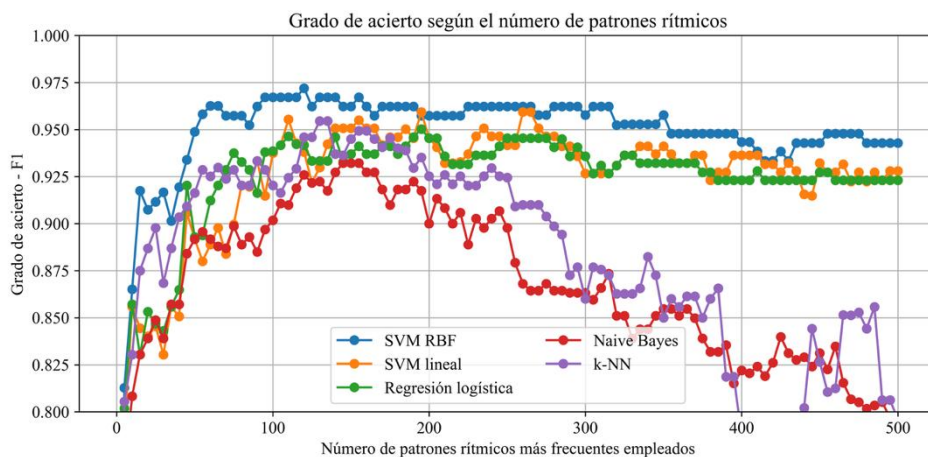


Figura 1. Grado de acierto (F1) de cinco clasificadores en la tarea binaria Lope/no Lope según el número de patrones rítmicos más frecuentes empleados.

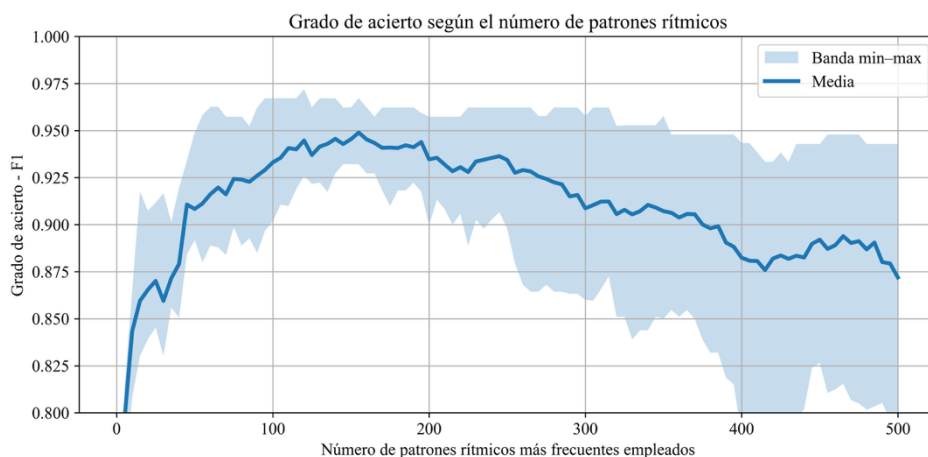


Figura 2. Media y rango (min-max) del grado de acierto (F1) entre los cinco clasificadores según el número de patrones rítmicos más frecuentes empleados.

Un aspecto especialmente informativo es que los máximos de rendimiento tienden a concentrarse cuando el análisis se restringe a un rango relativamente acotado de 100–200 patrones rítmicos. Esta concentración refuerza la hipótesis de que la señal autoral rítmica se articula principalmente en un núcleo de patrones frecuentes y que la incorporación de patrones menos comunes no garantiza mejoras y puede

introducir ruido. Este comportamiento contrasta con el patrón más habitual en estilometría léxica, donde el rendimiento suele estabilizarse a partir de un determinado número de *most frequent words* y se mantiene relativamente constante al ampliar el conjunto de rasgos (Cuéllar, 2024). En el caso rítmico, en cambio, los resultados sugieren la existencia de una ventana operativa: un intervalo de MFR en el que la señal es suficientemente fuerte y estable, pero fuera del cual el aumento de rasgos no aporta necesariamente capacidad discriminativa adicional, probablemente porque se incorporan patrones raros más sensibles a variación métrica, irregularidades textuales o ruido de escansión.

En definitiva, los resultados globales son muy satisfactorios: bajo las mejores configuraciones se alcanza un rendimiento cercano a la clasificación perfecta (F1 en torno a 0,95), lo que indica que, en un entorno de control, la señal rítmica permite discriminar con alta fiabilidad entre Lope y el resto. Precisamente por ello, resulta metodológicamente pertinente desplazar el foco desde el rendimiento agregado hacia los errores: ¿en qué textos falla de manera sistemática? ¿Qué comedias consideradas seguras dentro del bloque lopesco tienden a no ser reconocidas como tales? ¿Qué obras del conjunto no lopesco aparecen recurrentemente clasificadas como Lope? Y, sobre todo, ¿pueden identificarse razones formales o textuales que expliquen estos comportamientos? Para responder a estas preguntas, se realizó un análisis específico de los casos problemáticos. El análisis de errores revela un patrón nítido: las clasificaciones incorrectas no se distribuyen al azar, sino que se concentran de manera persistente en un conjunto muy reducido de obras, que tienden a comportarse como casos atípicos con independencia del clasificador y del umbral de MFR empleado. Esta estabilidad sugiere que, más que fluctuaciones del algoritmo, estamos ante textos cuya representación rítmica se aparta sistemáticamente del perfil esperado. En concreto, no se observa un sesgo simétrico hacia los falsos positivos: ninguna comedia del bloque no lopesco fue atribuida repetidamente a Lope de manera consistente; en cambio, seis comedias seguras de Lope fueron clasificadas de forma reiterada como no lopescas: *Adonis y Venus*, *El castigo sin venganza*, *El mejor alcalde, el rey*, *Fuenteovejuna*, *La reina Juana de Nápoles* y *Las bizzarrías de Belisa*. Conviene, por tanto, detenerse en estos textos para interrogar las razones de su comportamiento anómalo y proponer explicaciones plausibles.

Una primera constatación resulta especialmente llamativa desde el punto de vista filológico. Entre las seis obras que el modelo expulsa del

perfil rítmico lopesco figuran dos comedias con autógrafo conservado: *El castigo sin venganza* y *Las bazarrias de Belisa* (Presotto, 2000). En estos casos, la etiqueta de autoría no depende de inferencias externas o de tradición bibliográfica, sino de un testimonio material directamente vinculado al autor. Que el clasificador rítmico las sitúe fuera del bloque Lope obliga a interpretar el fenómeno como una limitación del método o de su representación, no como un indicio razonable de autoría alternativa. En términos metodológicos, estos falsos negativos refuerzan la idea de que la señal rítmica puede verse afectada por condiciones formales que no se capturan bien en una representación agregada a nivel de obra, incluso cuando la autoría es indiscutible por criterios externos. Cabe añadir, además, un posible factor de naturaleza cronológica. Ambas obras pertenecen a la etapa final de la producción de Lope (1631 y 1634), un periodo menos representado en el corpus en comparación con otras franjas cronológicas y con el repertorio de otros dramaturgos. Esta distribución desigual podría introducir un efecto de desajuste en la señal aprendida por el modelo, al dificultar la identificación de patrones característicos de esa fase tardía.

Algo similar ocurre con *El mejor alcalde, el rey*. No se conserva autógrafo para esta pieza en los repertorios consultados, pero su estatuto dentro del canon lopesco es estable y, además, supera sin dificultad los contrastes de estilometría léxica empleados en trabajos previos. En consecuencia, su clasificación reiterada como no Lope no puede leerse como una impugnación autorial, sino como una señal de que su perfil rítmico se aparta del comportamiento dominante que el modelo aprende como característico de Lope dentro del corpus de control. Esa desviación, sin embargo, se diluye en la agregación de patrones y no resulta fácilmente interpretable mediante una lectura cercana, ni siquiera para el filólogo experto, del mismo modo que ocurre con las distribuciones léxicas en estilometría. En la medida en que estos casos carecen, por ahora, de una explicación filológica específica y bien documentada, constituyen un recordatorio de que la estilometría rítmica, aun funcionando bien en promedio en un marco binario, no ofrece una estabilidad homogénea para todo el repertorio, y puede fallar de manera sistemática en piezas canónicas.

Los tres casos restantes permiten, en cambio, articular hipótesis más concretas sobre las condiciones que debilitan la señal rítmica. *La reina Juana de Nápoles* es un caso especialmente informativo porque combina una tradición de autenticidad relativamente asentada con indicios

modernos de anomalía. Morley y Bruerton (1968) no detectan disidencias estróficas y proponen una datación compatible con el repertorio lopesco; y, aunque algunos estudiosos ya habían emitido juicios negativos sobre la calidad de la pieza, ello no había desembocado en una impugnación firme de su autoría. Sin embargo, tanto Vega García-Luengos (2021) como Cuéllar (2022) coinciden en que la estilometría léxica no favorece su adscripción a Lope y, como mínimo, problematiza su integración en el corpus. En ese contexto, Vega García-Luengos (2021) reúne y ordena varias hipótesis explicativas, entre las que destaca una de carácter editorial: la intervención de agentes en la confección de las *Partes* y la posibilidad de reajustes o manipulaciones motivadas por criterios comerciales que habrían podido facilitar la inserción —voluntaria o no— de una obra no auténtica bajo marca lopesca. En *Adonis y Venus*, la explicación más plausible no remite tanto a una duda autorial como a un perfil formal excepcional. De hecho, en el experimento de cronología de Cuéllar (2023), la obra es la que arroja los resultados más negativos, lo que sugiere que presenta rasgos especialmente marcados que la separan del comportamiento medio aprendido por el modelo. Morley y Bruerton (1968) destacan que la pieza exhibe una variedad versificatoria inusualmente alta, con once formas métricas, y subrayan su brevedad relativa, que asocian al género, hasta caracterizarla como un posible libreto para una ópera. Estos rasgos convierten la comedia en un caso de polimetría intensa y de distribución métrica singular, condiciones que pueden traducirse en una señal rítmica menos estable cuando se trabaja con representaciones agregadas por frecuencias de patrones. Desde el punto de vista del experimento, este comportamiento es relevante porque indica que ciertas configuraciones métricas pueden situar una obra lopesca fuera del rango de variación que el modelo aprende como característico del autor. Por último, *Fuenteovejuna* constituye un caso especialmente sensible para el argumento metodológico del artículo porque su comportamiento anómalo no es exclusivo del plano rítmico. En trabajos previos se ha señalado que la estilometría léxica puede presentar tensiones internas en esta obra, en particular en la tercera jornada, lo que sugiere la conveniencia de tratarla como un texto estructuralmente heterogéneo a efectos de atribución y de perfilado estilométrico (Cuéllar, 2022). Que el análisis rítmico la clasifique de manera recurrente como no lopesca no debe interpretarse como un veredicto, pero sí como una convergencia de indicios que apunta hacia la necesidad de segmentación y control de heterogeneidad interna.

En conjunto, estos seis falsos negativos delimitan con precisión el espacio en el que la estilometría rítmica resulta robusta y, al mismo tiempo, los contextos en los que su señal puede debilitarse por factores métricos, textuales o de transmisión, incluso cuando la autoría no se discute en la tradición.

FUTURO Y CONCLUSIONES

Los resultados del experimento son, en conjunto, muy positivos. En un escenario binario Lope/no Lope y trabajando únicamente con información rítmica —esto es, con distribuciones de patrones acentuales extraídos automáticamente—, el sistema alcanza valores de F1 cercanos a 0,95 en las mejores configuraciones. Esta cifra es particularmente relevante porque se obtiene sin recurrir a léxico, temas, onomástica o rasgos semánticos, sino desde una dimensión formal independiente que, por definición, captura otro estrato de regularidad textual. En este sentido, la elevada eficacia de la estilometría léxica, que en muchos corpus roza el techo de rendimiento, tiende a eclipsar el alcance metodológico de un resultado rítmico que, por sí mismo, confirma que la prosodia computacional aporta señal autorial aprovechable y cuantificable.

El examen sistemático de clasificadores y umbrales de MFR confirma además un patrón consistente: la señal se concentra en un intervalo relativamente acotado de patrones frecuentes, típicamente en el entorno entre 100 y 200. Mientras que en estilometría léxica es habitual observar una estabilización del rendimiento a partir de un número suficiente de palabras frecuentes, aquí los máximos se agrupan en una ventana operativa más estrecha, lo que sugiere que la inclusión de patrones menos frecuentes no añade necesariamente información discriminativa y puede introducir ruido asociado a polimetría, irregularidades o inestabilidad de la escansión. En términos prácticos, este resultado ofrece una guía de parametrización: el ritmo parece funcionar mejor cuando se explota un núcleo recurrente de hábitos prosódicos, y no cuando se amplía indefinidamente el espacio de rasgos.

Ahora bien, estos mismos experimentos obligan a mantener una prudencia interpretativa estricta. Incluso en la configuración de mejor rendimiento, persisten errores que, por su naturaleza, resultan metodológicamente inaceptables si se pretendiera trasladar el procedimiento de forma directa a obras problemáticas: que dos comedias con autógrafo se clasifiquen sistemáticamente como no lopesca muestra

que, pese al buen comportamiento global, la señal rítmica todavía puede colapsar en casos particulares sin que ello refleje una conclusión autorial razonable. Este hallazgo delimita con claridad el estado actual del enfoque: la estilometría rítmica es ya operativa y prometedora como evidencia complementaria en tareas binarias controladas, pero no puede presentarse todavía como una herramienta autónoma en escenarios donde la interpretación dependa de un pequeño número de obras atípicas.

De cara al trabajo futuro, el avance debe desplegarse en dos direcciones complementarias. La primera concierne a la transparencia y explicabilidad del proceso de clasificación. Parte de la utilidad del enfoque depende de poder reconstruir por qué una obra concreta queda a uno u otro lado de la frontera, y de distinguir entre decisiones sostenidas por un núcleo de rasgos estables y decisiones inducidas por ruido o por configuraciones métricas singulares. Esto sugiere priorizar modelos más interpretables o, al menos, incorporar procedimientos de explicación posterior que permitan inspeccionar la contribución relativa de los patrones rítmicos en cada predicción, especialmente en los *outliers* reiterados. En esta misma línea, será pertinente en futuras iteraciones del experimento explorar con mayor detalle el control de la distribución cronológica del corpus, en particular en lo que respecta al equilibrio entre obras tardías de Lope y del conjunto no loresco, con el fin de evaluar hasta qué punto estos factores pueden influir en la estabilidad de la señal rítmica. La segunda dirección concierne al enriquecimiento de la propia representación rítmica. En el estado actual, el sistema opera sobre un conteo agregado de patrones; este esquema, aunque eficaz, simplifica en exceso una señal que es secuencial, jerárquica y condicionada por la polimetría. En consecuencia, parece necesario introducir representaciones que preserven estructura: segmentación por jornadas o bloques métricos, modelado condicionado por metro, análisis de transiciones entre patrones, o tratamiento explícito de la variación interna de la obra. En ese marco, una línea particularmente fértil consiste en aproximar el ritmo como señal y adoptar herramientas procedentes del análisis de señales para captar regularidades que se pierden en el conteo global. Técnicas espectrales como la transformada de Fourier permiten reexpresar una señal en términos de componentes periódicas y podrían contribuir a identificar firmas prosódicas más legibles y robustas frente a ruido local, al tiempo que facilitan comparaciones entre obras con distribuciones superficiales similares, pero organización interna distinta. Este tipo de enfoques no implica abandonar el nivel rítmico, sino refinarlo: pasar de un inventario

estadístico de patrones a descriptores que modelen periodicidad, densidad de alternancias y estabilidad a distintas escalas, con el objetivo explícito de reducir falsos negativos en textos canónicos y de aumentar la confianza cuando el método se aplique a atribuciones controvertidas.

En síntesis, el trabajo confirma que la estilometría rítmica, formulada como problema binario y con una selección adecuada de patrones frecuentes, ofrece resultados muy competitivos y aporta evidencia independiente para la atribución autorial. Al mismo tiempo, el análisis de los errores muestra con claridad dónde se concentran sus límites actuales y qué aspectos requieren desarrollo: mejorar la explicabilidad de las decisiones y enriquecer la modelización del ritmo para que la señal sea más estable en obras formalmente singulares o internamente heterogéneas. En la medida en que sigan ampliándose los corpus escandidos y se consoliden procedimientos de control de calidad y de representación más estructural, el ritmo puede convertirse en un componente cada vez más fiable dentro de un marco integrador de evidencias autoriales.

BIBLIOGRAFÍA

Antonucci, Fausta (2020). “La polimetría en *La dama boba*: funciones poéticas y dramáticas”. En Javier Espejo Surós y Carlos Mata Induráin (eds.). *Preludio a «La dama boba» de Lope de Vega (historia y crítica)*. Pamplona: Servicio de Publicaciones de la Universidad de Navarra, pp. 111-129.

Artelope (2011-2026). Base de datos y argumentos del Teatro de Lope de Vega. Joan Oleza et al. (dirs.). Universitat de València, <https://artelope.uv.es> [10/01/2026].

Cuéllar, Álvaro (2022). *Propuestas de autoría y datación en el teatro de Lope de Vega a la luz de la Inteligencia Artificial*. Tesis doctoral (PhD). University of Kentucky. DOI: <https://doi.org/10.13023/etd.2022.428>.

Cuéllar, Álvaro (2023). “Cronología y estilometría: datación automática de comedias de Lope de Vega”. *Anuario Lope de Vega. Texto, literatura, cultura*, 29, pp. 97-130. DOI: <https://doi.org/10.5565/rev/anuariolopedevega.483>.

- Cuéllar, Álvaro (2024). “Stylometry and Spanish Golden Age Theatre: An Evaluation of Authorship Attribution in a Control Group of One Hundred Undisputed Plays”. En Robert Hesselbach, Juan Calvo Tello, Ulrike Henny-Krahmer, Christof Schöch y Daniel Schlör (eds.). *Digital Stylistics in Romance Studies and Beyond*. Heidelberg: Heidelberg University Publishing, pp. 101-117. DOI: <https://doi.org/10.17885/heiup.1157.c19368>.
- Cuéllar, Álvaro (2025a). “From Atoms to Waves: Rhythmic Stylometry for Authorship Studies of Early Modern Spanish Theatre”. *Janus. Estudios sobre el Siglo de Oro*, 14. DOI: <https://doi.org/10.17979/janus.2025.14.12653>.
- Cuéllar, Álvaro (2025b). *Rhythmic Corpus of 100 Spanish Golden Age Plays* [dataset]. Zenodo. DOI: <https://doi.org/10.5281/zenodo.17537343>.
- Kroll, Simon (2022). *Sonido y afecto en Calderón. Un estudio de las asonancias*. Kassel: Edition Reichenberger.
- Kroll, Simon y Fernando Sanz-Lázaro (2022). “Romances teatrales entre Mira de Amescua, Calderón y Lope: ritmo, asonancia y cuestiones de autoría”. *Revista de Humanidades Digitales*, 7.1, pp. 1-18. DOI: <https://doi.org/10.5944/rhd.vol.7.2022.31620>.
- Kroll, Simon y Fernando Sanz-Lázaro (2023). “Ritmo e inteligencia artificial: nuevas perspectivas sobre teatro lopesco desde las humanidades digitales”. *Anuario Lope de Vega. Texto, literatura, cultura*, 29, pp. 351-375. DOI: <https://doi.org/10.5565/rev/anuariolopedevega.491>.
- Kroll, Simon y Fernando Sanz-Lázaro (2024). *Versos y estructuras teatrales áureos: A corpus of Spanish plays with dramatic and metric information*. PHAIDRA (Universidad de Viena). DOI: <https://doi.org/10.25365/phaidra.471>.
- Marco, Guillermo, Javier de la Rosa, Julio Gonzalo, Salvador Ros y Elena González-Blanco (2021). “Automated metric analysis of Spanish

- poetry: Two complementary approaches”. IEEE Access, 9, pp. 51734-51746. DOI: <https://doi.org/10.1109/ACCESS.2021.3069635>.
- Marco, Guillermo y Julio Gonzalo (2021). “Escansión automática de poesía española sin silabación”. *Procesamiento del Lenguaje Natural*, 66, pp. 77-87. DOI: <https://doi.org/10.26342/2021-66-6>.
- Morley, S. Griswold y Courtney Bruerton (1968). *Cronología de las comedias de Lope de Vega. Con un examen de las atribuciones dudosas, basado todo ello en un estudio de versificación estrófica*. Madrid: Gredos.
- Navarro-Colorado, Borja (2018). “A metrical scansion system for fixed-metre Spanish poetry”. *Digital Scholarship in the Humanities*, 33.1, pp. 112-127. DOI: <https://doi.org/10.1093/llc/fqx009>.
- Pedregosa, Fabian et al. (2011). “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research*, 12, pp. 2825-2830.
- Presotto, Marco (2000). *Le commedie autografe di Lope de Vega: catalogo e studio*. Kassel: Reichenberger.
- Sanz-Lázaro, Fernando (2022). “Planteamientos digitales del drama aurisecular: automatización del análisis métrico de obras teatrales”. En Íñigo González Cabeza, Érica Redruello Vidal y Raquel de la Varga Llamazares (eds.). *La escritura en el tiempo: Investigaciones en torno a la literatura hispánica*. León: Universidad de León, pp. 109-128.
- Sanz-Lázaro, Fernando (2023a). *Ritmo y estructura de la comedia áurea: Posibilidades y límites del análisis digital automático del teatro español clásico* (tesis doctoral, Universität Wien). DOI: <https://doi.org/10.25365/thesis.75102>.
- Sanz-Lázaro, Fernando (2023b). “Del fonema al verso: una caja de herramientas digitales de escansión teatral”. *Revista de Humanidades Digitales*, 8, pp. 74-89. DOI: <https://doi.org/10.5944/rhd.vol.8.2023.37830>.

Sanz-Lázaro, Fernando (2024). “libEscansión: A recursive precedence approach to metrical scansion”. *Digital Humanities Quarterly*, 18.3, <https://github.com/fsanzl/libescansion> [10/03/2026].

Vega García-Luengos, Germán (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. DOI: <https://doi.org/10.5209/tret.74625>.