

Aproximación al análisis cuantitativo de la reutilización textual en el teatro del Siglo de Oro. El caso de las colaboradas de Agustín Moreto

Quantitative Analysis of Textual Reuse in the Theatre of the Spanish Golden Age: The Case of the Collaborative Plays of Agustín Moreto

José Luis Losada Palenzuela

<https://orcid.org/0000-0002-6530-1328>

Universidad de Wrocław

POLONIA

jose-luis.losada@uwr.edu.pl

[*Hipogrifo*, (issn: 2328-1308), 12.1, 2024, pp. 143-158]

Recibido: 30-01-2024 / Aceptado: 07-03-2024

DOI: <http://dx.doi.org/10.13035/H.2024.12.01.10>

Resumen. La especificidad del contexto sociocultural del teatro español del siglo xvii favoreció la práctica de la reutilización textual. Los autores se servían de obras ya escritas o representadas, construyendo una práctica intertextual mixta basada en la refundición, la adaptación y la reutilización (entre otras definiciones). El estudio de la reutilización sigue siendo un elemento importante de la crítica textual para establecer relaciones textuales. Dentro del corpus teatral, las comedias colaboradas suponen un subconjunto muy representativo del fenómeno, pues es-

Este trabajo ha surgido dentro de los proyectos de investigación «El teatro áureo en colaboración: textos, autorías, ámbitos literarios de sociabilidad y nuevos instrumentos de investigación (tac)», PID2020-117749GB-C22 y «Computational Text Reuse Detection in Literary Texts», BPIDUB.29.2022.

tas raras veces ofrecían contenidos originales. Este trabajo presenta los resultados de los métodos cuantitativos (un conjunto de algoritmos ofrecidos por TRACER) para detectar automáticamente la reutilización en un subcorpus teatral de comedias colaboradas participadas por Agustín Moreto. Se muestra que no solo es posible identificar automáticamente secuencias repetidas, sino que una clasificación general refleja cierta filiación entre las colaboradas basada en la repetición de automatismos, típicos de la creación teatral.

Palabras clave. Reutilización textual, comedias colaboradas; Agustín Moreto; análisis textual cuantitativo; TRACER.

Abstract. The specificity of the sociocultural context of Spanish Drama in the 17th century favoured the practice of textual reuse. Authors made use of plays already written or performed, constructing a mixed intertextual practice based on recasting, adaptation and reuse (among other definitions). The study of repeated sequences remains an important element of textual criticism to establish textual relationships. Within the Spanish drama corpus, comedies co-written by one or more authors represent a very significant subset of the phenomenon, as they rarely offer original content. This paper presents the results of quantitative methods (a set of algorithms offered by TRACER) to automatically detect reuse in a theatrical subcorpus of co-written plays. We show that not only is it possible to automatically identify sequences of parallel passages, but that a general classification reflects a certain filiation between co-written plays based on the repetition of automatism, typical of drama creation.

Keywords. Text reuse; Co-written drama; Agustín Moreto; Quantitative text analysis; TRACER.

INTRODUCCIÓN

El teatro español del siglo xvii se caracterizó por la práctica extendida de la reutilización textual, favorecida por el contexto sociocultural de la época. Los autores recurrían a obras ya escritas o representadas, construyendo una práctica intertextual basada en la refundición, adaptación o reutilización. Este fenómeno es un elemento clave para la crítica textual, ya que permite establecer relaciones y conexiones entre las diferentes obras. Un ejemplo representativo de esta práctica son las comedias colaboradas, que rara vez ofrecían contenidos completamente originales, sino que se basaban en la reelaboración de materiales previos.

En este trabajo presentamos las posibilidades que ofrece el análisis textual cuantitativo para detectar de forma automática la reutilización de versos, poniendo el foco en las comedias colaboradas en las que participa Agustín Moreto.

Antes de abordar este asunto, me gustaría exponer la distinción entre una expresión (o cadena de texto) conocida y una desconocida, porque implica planteamientos distintos en la identificación de la reutilización.

Si nos interesa localizar una expresión de partida conocida (una frase, verso, una colocación, etc.), siempre podemos comprobar su presencia o su distribución en un corpus. Es una práctica común. La efectividad de los resultados depende del diseño de los corpus teatrales, ya sean bases de datos en línea (TESO, TEXORO, CORDE), repositorios con accesos a través de una interfaz de programación (DRACOR) o meros listados de archivos en nuestro ordenador.

Existen tecnologías asequibles para transformar simples listados de textos en una base de datos que permita, en una primera instancia, realizar búsquedas de patrones conocidos o analizar, por ejemplo, concordancias, colocaciones, análisis estadísticos, etc.¹.

Keyword-in-context with 4 matches.

```
[text6, 6125:6128] salte fuera, | veneno en taza dorada | , sepulcro hermoso
[text7, 5840:5843] liga prende; | veneno en taza dorada | que con resplandor
[text12, 10373:10376] los hombres, | veneno en taza dorada | , codicia de
[text18, 11649:11652] salte fuera, | veneno en taza dorada | , sepulcro hermoso
```

Lexical dispersion plot



Fig. 1. KWIC y dispersión de la expresión «veneno en taza dorada» (Lope, Tirso y Calderón)

Por otro lado, podríamos querer localizar cadenas de texto desconocidas que suponemos que se repiten en uno o más lugares. En este caso tenemos dos posibilidades: la primera, detectar la presencia de una fuente en un destino (simple o múltiple), es decir, encontrar los pasajes citados de un texto fuente en otro texto o en un corpus, por ejemplo, la Biblia en el corpus teatral. En la segunda posibilidad, la fuente y el destino son intercambiables: comparamos todos los textos entre sí para detectar los emparejamientos de reutilización, los pasajes paralelos, por ejemplo, en un corpus de obras de teatro (fuente y destino pueden ser también un solo documento si buscamos reutilizaciones dentro de la misma obra).

1. Este tipo de análisis textual cuantitativo comparte metodología con la Lingüística de corpus. Los lenguajes de programación (R o Python), así como el *software* específico (Antconc, Sketch Engine, #Lancs-Box X, etc.) permiten fácilmente abordar este tipo de consultas, si se dispone de los textos. La Fig. 1 se ha realizado en R usando el paquete *quanteda*.

En este trabajo abordamos esta segunda posibilidad: encontrar automáticamente reutilizaciones en un corpus de comedias cuando desconocemos la cadena de texto. Se trata de recuperar automáticamente versos, partes de versos, tiradas iguales o casi iguales, etc., compartidos por dos o más obras.

El objetivo es evaluar el siempre discutido grado de originalidad de las colaboradas, así como contribuir a los estudios sobre reescritura teatral, que han sido desde su inicio un método de apoyo para establecer la filiación entre textos, la atribución autorial o la colaboración.

Al comienzo expondré el concepto de reutilización y las implicaciones de la exploración cuantitativa. Después contextualizaré la metodología, el corpus y la configuración de los parámetros; haré una descripción de los resultados y cerraré con una reflexión sobre las posibilidades de este método. Veremos que de esta manera podemos extraer no solo reutilizaciones de un corpus teatral, sino también, en cierta medida, individualizar alguna característica específica de las comedias colaboradas.

EL CONCEPTO DE REUTILIZACIÓN

La reutilización consiste en una repetición significativa de un texto. Es un concepto amplio, presente en varios contextos disciplinares, pues su sentido salta los márgenes de la mera repetición. Es un fenómeno clásico de las Humanidades debido a su dependencia textual.

La génesis del corpus teatral español pasa, literalmente, por muchas manos. Los autores aprovechan repetidas veces textos propios y ajenos en su creación dramática. La conocida delación de Jerónimo de Cáncer sobre las artes plagiarias de Moreto, quien hallaba en comedias viejas una mina de inspiración (Sáez Raposo, 2010, p. 200), dejaba intuir un tipo de proceder, que ciertamente no se limita a un solo autor.

No contamos, sin embargo, con una taxonomía unitaria de la reutilización. A la primera taxonomía teatral de Ruano de la Haza (1998, p. 36) con términos como refundición, reelaboración, reconstrucción, adaptación, reutilización, la acompañan otras reflexiones más genéricas sobre inter e intratextualidad, autoplagio, reciclaje (Sánchez Jiménez, 2014; Sáez, 2013), auto-reescritura, hetero-reescritura (Vitse, 1998; Vega García-Luengos, 1998) o versión (Pinillos, 2002), casi siempre en el marco de trabajos individuales sobre obras y autores (Vila Carneiro, 2017; Fernández Mosquera, 2015; Sáez Raposo, 2010; Escudero Baztán, 2013; Rodríguez Gallego, 2010).

A pesar de la precaución en el uso de la intertextualidad para la filiación², sigue siendo evidente la utilidad filológica de individuar expresiones o pasajes paralelos para la filiación de textos y autores (Vega García-Luengos, 2017 y 2021), siempre que concurren otros métodos como la bibliografía material, versificación, ortología o, más recientemente, la estilometría.

Por otra parte, la identificación automática de unidades textuales es un ámbito de estudio de las ciencias computacionales, de los llamados sistemas de recuperación de información, que permiten, dados ciertos términos de consulta, extraer automáticamente los documentos más relevantes. Muchas de las técnicas del Procesamiento del Lenguaje Natural y de la Semántica Distribucional forman parte de este ámbito, donde la reutilización interesa en la eliminación de la redundancia, la descontaminación de textos, la creación de corpus, la detección del plagio o la republicación en el periodismo histórico³.

METODOLOGÍA

Para poder aplicar métodos cuantitativos digitales a conceptos literarios es necesario convertir estos en unidades cuantificables, es decir, operacionalizarlos⁴. Quiero servirme aquí del concepto de reutilización combinando la clásica definición de Ruano con una definición más computacional, en la que la reutilización relaciona un documento de partida y uno de llegada que, estadísticamente hablando, comparten un número significativo de palabras:

la "reutilización" consiste en el aprovechamiento de una escena o pasaje textual en diferentes piezas teatrales [...], pero también se podrían incluir en esta categoría las canciones, versos y otros tipos de textos literarios injertados en una comedia con fines diferentes a los que tenían originalmente (Ruano de la Haza, 1998, p. 36).

Text re-use concerns the recycling of complex textual information which contains a statistically significant number of words in common between a source document and a target document (Büchler *et al.*, 2014, p. 225).

2. «Lo que hoy llamaríamos plagio y autoplagio —y que en la época se conocía por el menos litigioso término de imitación— era tan común en el teatro áureo que es simplemente imposible demostrar la descendencia o ascendencia de una comedia sobre otra apoyándose en coincidencias argumentales, estilísticas o léxicas. El único método objetivo de relacionar diferentes ediciones o versiones de una misma comedia es el de la crítica textual» (Ruano de la Haza, 1992, pp. 30-32).

3. El tema es muy amplio y forma parte esencial del espacio textual digital. Véase en español, a modo de breve introducción, Navarro Colorado (2022) para ejemplos sobre *word embeddings*, *TF/IDF*, o *topic modeling*.

4. Para el concepto de operacionalización aplicado a los estudios literarios véase Pichler y Reiter (2022) y Moretti (2013), quien fue el primero en apuntar la necesidad de diseñar el paso entre metodologías, precisamente porque «most literary concepts are emphatically not designed to be quantified» (Moretti, 2013, p. 114).

De forma muy simplificada, el procedimiento se reduce a la computación de frecuencias de unidades léxicas (*tokens*) y a la medida de distancias entre ellas. Algunos de los algoritmos usuales vienen ya implementados en las librerías de los lenguajes de programación o en *software* como TRACER, *passim*, *textreuse*, por tanto, los aspectos de programación más complejos ya están resueltos, pero otros muchos parámetros y decisiones entran en juego antes de su aplicación: ¿comparamos párrafos, oraciones, versos?, ¿bigramas, trigramas, palabras?, ¿eliminamos las palabras función?, sin mencionar la necesaria preparación previa de un corpus digital.

Me voy a servir en este caso de TRACER (*Text Reuse Detection Machine*), un conjunto de algoritmos desarrollados en Java, diseñado para trabajar con textos históricos independientemente de la lengua. Ha sido ya usado para analizar distintas traducciones inglesas de la Biblia o para evaluar la reutilización textual en autores clásicos griegos (Büchler *et al.*, 2014; Büchler *et al.*, 2013).

CORPUS

He reunido un corpus de unas 500 obras (extraído de fuentes digitales con textos normalizados: Canon60, BCV, AHCT, ARTELOPE, DRACOR, Moretianos), de las que he seleccionado un subcorpus de comedias en colaboración con Agustín Moreto como nexo de unión en todas ellas.

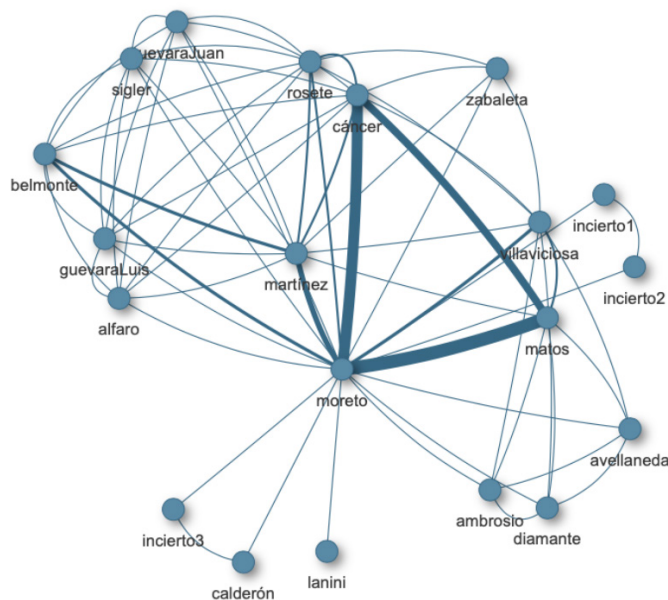


Fig. 2. Red de autores en colaboradas, con Moreto como nexo de unión⁵

5. El grafo se puede consultar en línea de forma interactiva en https://editio.github.io/presentations/imagenes/cola_moreto_colaboradas. Los datos se basan en la información disponible en moretianos.com.

Las colaboradas representan un buen banco de trabajo, ya que por definición sus temas rara vez son originales: «Il soggetto dei componimenti drammatici scritti *de consuno* non è mai originale» (Alviti, 2006, p. 17). Se trata de refundiciones, de reelaboraciones de otras obras, para adaptarlas a nuevos públicos o nuevos contextos. Aunque las implicaciones son, por supuesto, mucho más profundas, cabe imaginar que puede ser más fácil producir una obra en colaboración si se tiene un modelo a seguir. Así pues, el corpus cuenta con obras de teatro en colaboración y posibles modelos previos: Tirso de Molina (14 %), Lope de Vega (56 %), Pedro Calderón de la Barca (18 %), Agustín de Moreto (8 %), Agustín de Moreto en colaboradas (4 %).

PREPROCESO Y PARÁMETROS

Los datos de entrada estructurados de todo el corpus, antes de cualquier cómputo, necesitan una estructura específica (Fig. 3). Se han eliminado todos los paratextos (*dramatis personae*, disdascalias, etc.), se ha restablecido de forma automática (siempre que ha sido posible) en una sola línea la habitual división tipográfica de los versos partidos, se ha suprimido la puntuación y los signos diacríticos, estableciendo la unidad de comparación en el verso.

reuse_units	id	text_verse	play
1	0100001	hipogrifo violento	calderon_LaVidaEsSueno_CANON
2	0100002	que corriste parejas con el viento	calderon_LaVidaEsSueno_CANON
3	0100003	donde rayo sin llama	calderon_LaVidaEsSueno_CANON
4	0100004	pajaro sin matiz pez sin escama	calderon_LaVidaEsSueno_CANON
...			
...			
1 409 516	50904166	yo soy si acaso os agrada	tirso_villanaVallecas_COMED
1 409 517	50904167	la villana de vallecas	tirso_villanaVallecas_COMED
1 409 518	50904168	mas si no no sere nada	tirso_villana_vallecas_COMED

Fig. 3. Tabla de unidades de comparación

El resultado es una tabla con aproximadamente un millón y medio de unidades de reutilización, donde la variable *id* identifica las obras y las unidades por obra. TRACER usa un algoritmo basado en el enlace de elementos: no compara todo con todo (*Brute Force*), sino que solo comprueba los elementos comunes en las unidades de reutilización (*Feature Based Linking*). En nuestro caso hemos elegido los versos, pero podrían ser oraciones, párrafos, etc. Esta es ya una primera tokenización. Estas unidades, a su vez, se vuelven a segmentar, pasan por una segunda tokenización en elementos (*features*): en el *featuring* deberemos decidir si los elementos son palabras, bigramas, etc. La razón es optimizar la computación para reducir el número total de comparaciones sin perder precisión en los resultados. A partir de aquí sigue una lista de decisiones sobre, por ejemplo, si se eliminan las

palabras función (con una frecuencia muy alta y, por tanto, se repiten a menudo) o, al contrario, eliminamos elementos que aparecen poco o una sola vez (hápx), con lo que no tiene sentido incluirlos en la comparación.

Además, esto es importante, se debe determinar cuál es el porcentaje de elementos comunes en una unidad para que sea calificada como reutilización (*scoring*): cuanto mayor sea el porcentaje fijado, mayor deberá ser la semejanza de las unidades. Esto nos permite calificar como reutilización una unidad, aunque solo coincida parcialmente, de esta manera podemos acceder no solo a reutilizaciones literales, sino a reutilizaciones casi literales, dependiendo del porcentaje que usemos.

Recuérdese la famosa décima compartida entre *El príncipe constante* y *Amor, honor y poder* de Pedro Calderón de la Barca, «Lisonjera, libre, ingrata / [...] / ingrata porque corría». En un ejercicio de autoplagio (Vila Carneiro, 2017, pp. 736-737), Calderón reutiliza la décima en dos obras, pero no exactamente igual, ya que algunos versos presentan ligeras variaciones (en cursiva): «lisonjera se *dilata*» frente a «lisonjera se *desata*»; «*hace* apacible corriente» frente a «*hizo* apacible corriente»; «y ingrata porque corría» frente a «ingrata porque corría». Fijando un *scoring* del 100 % no recuperaríamos estos versos como unidad de reutilización. Un porcentaje menor recuperaría quizá unidades menos interesantes, basadas solo, por ejemplo, en preposiciones o artículos compartidos en un verso. El porcentaje se ve afectado, por tanto, por otras decisiones en el *featuring*, por ejemplo, si eliminamos las palabras función (la variación en «(ø | y) ingrata porque corría» dejaría de existir), los signos diacríticos («qué», «más» frente a «que», «mas») o si usamos bigramas o palabras, etc.

Definir el conjunto de criterios pasa primero por una validación. En nuestro caso hemos realizado una validación mínima⁶ partiendo de dos ejemplos extremos de reutilización en Calderón de la Barca, «por antonomasia, el dramaturgo áureo de la reescritura» (Vitse, 1998, p. 6). El primero es el caso de la jornada completa compartida entre *Los cabellos de Absalón* de Calderón y *La venganza de Tamar* de Tirso de Molina (Sloman, 1958, p. 95). TRACER recupera sin problemas la jornada (también en sus variaciones) con los parámetros en: unidad de versos, sin diacríticos, sin palabras función, bigramas, enlace externo y *score* mínimo del 0.5.

El segundo es el ejemplo ya mencionado de la décima, con los siguientes parámetros: unidad en versos, sin diacríticos, con palabras función, bigramas, enlace externo y *score* mínimo del 0.5. Con esta configuración, los algoritmos recuperan todos los versos literales de la décima, pero dejan fuera dos de los tres versos con variación.

6. Evaluar el rendimiento de TRACER para este corpus usando medidas de precisión y exhaustividad supondría todo un experimento adicional. Sin embargo, existen trabajos de validación con TRACER para un corpus bíblico, que han servido de guía para nuestra investigación. Véase Franzini et al., 2018.

RESULTADOS

Hemos fijado para el análisis los parámetros en: unidad de versos, sin diacríticos, con palabras función, bigramas, enlace externo y *score* mínimo del 0.6. Como veremos, se trata de un compromiso que nos permita recuperar pasajes paralelos *verbatim* y casi *verbatim* sin añadir mucho ruido en los resultados.

n	score	obra.x	obra.y
1	1	que soy niño y solo y nunca en tal me vi	que soy niño y solo y nunca en tal me vi
2	1	pues si un día es el siglo de las flores	pues si un día es el siglo de las flores
3	1	señal que es noble el alma que esta en ella	señal que es noble el alma que esta en ella
4	1	mas para que lo que es tan cierto he puesto	mas para que lo que es tan cierto he puesto
5	1	que es muerto el rey entre el confuso y vario	que es muerto el rey entre el confuso y vario
...			
37 346	1	que es esto cielos que es esto	que es esto cielos que es esto
37 347	1	que es esto cielos que es esto	que es esto cielos que es esto
37 348	1	todos los cuatro elementos	todos los cuatro elementos
...			
237 344	0.6	en la frente de un traidor	en la frente de un aspid
237 345	0.6	en las aras de su temple	en las aras de su altar
237 346	0.6	que es esto cielos que escucho	que es esto cielos que es esto

Fig. 4. Tabla de resultados de reutilización totales

Obtenemos una tabla (Fig. 4) con un total de 237 346 reutilizaciones distribuidas en 504 obras (casi la mayoría de las obras muestran reutilización, por tanto). Tenemos varias maneras de proceder a la interpretación de los resultados.

Para una primera valoración general uso (ya fuera de TRACER) análisis de redes. Construyo un grafo con todas las obras que comparten reutilización marcando el peso ponderado de las aristas según el valor de cada unidad de comparación (*score*), es decir, cada nodo representa una obra que se conecta con otra si comparten reutilización. El tamaño de la arista (de la conexión) representa una mayor o menor reutilización entre las obras; el color del nodo diferencia a los autores. Para la espacialización (distribución general de los nodos) uso el algoritmo *GEM force-directed layout* (Jacomy *et al.*, 2014), que agrupa los nodos con el mayor número de conexiones separando visualmente asimismo los grupos entre ellos.

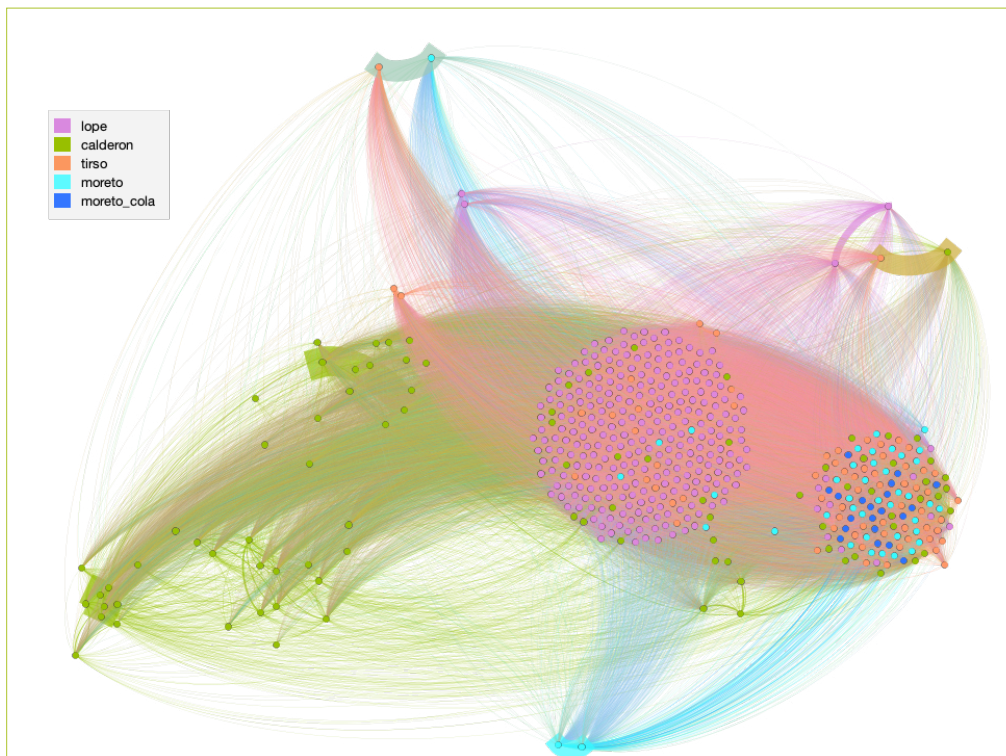


Fig. 5. Grafo de reutilización en el corpus

En un mínimo análisis descriptivo vemos que la gran densidad de aristas refleja un elevado número de reutilizaciones. El intercambio se agrupa en un clúster central ocupado mayoritariamente por Lope de Vega, un segundo clúster lateral ocupado por Moreto y las colaboradas participadas por él. Calderón aparece disperso, aunque un clúster periférico agrupa sobre todo sus Autos Sacramentales presentes en el corpus. Una primera intuición ya nos indica la importante semejanza de frecuencias de reutilización de las colaboradas con el propio Moreto.

Debemos tener en cuenta que dos nodos, dos obras, estarán conectadas con solo tener un verso en común, por ejemplo: «en la frente de un *traidor*» y «en la frente de un *áspid*»⁷, este sin necesidad además de que sea literal, como vemos (diferencias en cursiva). Traducido a la terminología que venimos usando, es suficiente con que el total de *score* que compartan sea 0.6, rango mínimo de reutilización.

Si, en cambio, construimos el grafo (Fig. 6) fijando como parámetro mínimo un *score* total de 25 (el rango mayor en este caso es 1080⁸), desaparecen el 99% de las aristas (de las reutilizaciones), quedando solo 90. Desaparecen asimismo del grafo

7. En los ejemplos de reutilización que menciono a partir de ahora se restituyen los signos diacríticos, pero téngase en cuenta que el análisis ha sido llevado a cabo sin ellos, es decir, «aspid», «senor», «lagrimas», etc.

8. A partir de ahora mostraré todos los *scores* con números enteros, sin decimales.

el 75 % de las obras, quedando 76 obras, del total de 504 con reutilización. Suprimir el grado de reutilización de la escala baja causa que permanezca visible solo el 1 % de aristas, esto es, la suma de la mayoría reutilizaciones entre dos obras es baja, uno o dos versos. Lope de Vega desaparece completamente, porque la mayoría de reutilizaciones se dan entre sus obras en un rango muy bajo. Esto no conlleva una valoración cualitativa: un pasaje de únicamente dos versos puede ser muy significativo analizado en un contexto particular.

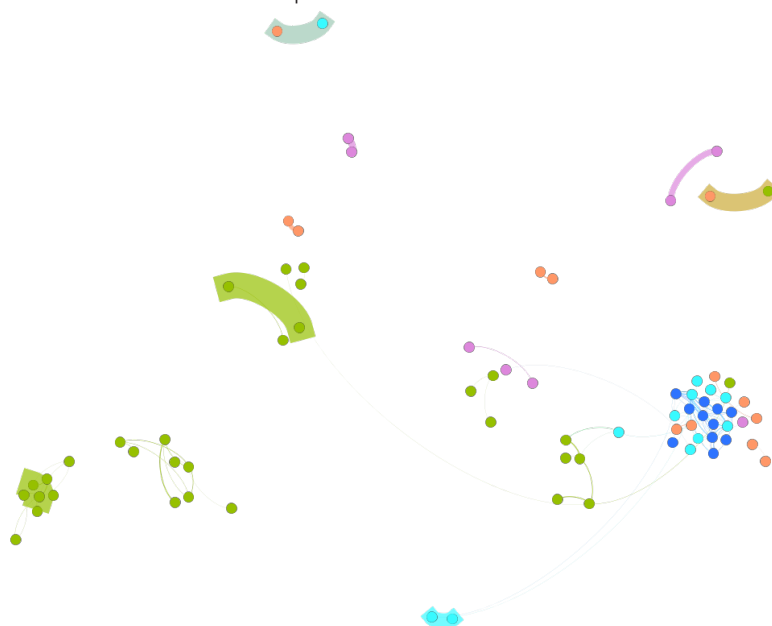


Fig. 6. Subgrafo de reutilización con el peso ponderado de aristas a partir de un score de 25^9

De esta manera emergen con claridad en el subgrafo, por ejemplo, las reutilizaciones evidentes en las refundiciones calderonianas, el saqueo a Tirso o una de las pocas reutilizaciones documentadas en Lope de Vega:

- score 1080: *Triunfar muriendo* y *La segunda esposa* (Calderón).
- score 1022: *El diablo mudo*, primera y segunda versión (Calderón).
- score 724: *La villana de Vallecas* (Tirso) y *La ocasión hace al ladrón* (Matos Frago) ¹⁰.

9. El subgrafo puede ser consultado de forma interactiva en https://editio.github.io/presentations/reuse_subset.

10. *La ocasión hace al ladrón* ha sido tradicionalmente atribuida a Moreto, por eso figura en este corpus. Elena Muñoz Rodríguez ha fijado el texto que atribuye ahora a Matos Frago con la estilometría. Véase www.moretianos.com, en obras atribuidas a Moreto.

- score 695: *La venganza de Tamar* (Tirso) y *Los cabellos de Absalón* (Calderón)¹¹.
- score 629: *El parecido* y *El parecido en la corte* (Moreto)¹².
- score 282: *Los pleitos de Ingalaterra* y *La Corona de Hungría* (Lope)¹³.
- score 254: *La dama boba* y *El villano en su rincón* (Lope)¹⁴.

Si volvemos la vista a las colaboradas, vemos que su clúster destaca por su protagonismo: a diferencia de lo que ocurrió con el clúster de Lope, más de la mitad se mantienen agrupadas entre sí y con Moreto. Tirso aparece también en el clúster por el número similar de conexiones, pero un análisis más cercano nos muestra que no comparte ninguna reutilización con las colaboradas. Dicho de otro modo, las colaboradas muestran un nivel conjunto de reutilización mayor y exclusivo con obras de Moreto, excepto una: *El príncipe perseguido* (Belmonte, Moreto, Martínez de Meneses), que reproduce fragmentos (score 26) de *El gran duque de Moscovia* y *Emperador perseguido* (Lope) (Baczyńska, 2019).

Si se inspecciona someramente el contenido de estas reutilizaciones en las colaboradas, nos damos cuenta de que no contienen una estrofa completa, canciones, tiradas duplicadas. Son más bien automatismos, expresiones típicas y conocidas, pero que se repiten a menudo entre las colaboradas y las comedias escritas en solitario por Moreto.

El hijo pródigo (Cáncer, Moreto y Matos Fragoso) comparte un score de 59 con *Hasta el fin nadie es dichoso* (Moreto), pero se trata de usos literales del tipo «ay de mí cielos qué miro», «válgame el cielo qué escucho» (muy recurrentes además a lo largo de las obras) o casi literales como «merezco en fe de mi amor» y «en fe de mi amor heroico». De manera similar, de nuevo *El hijo pródigo* (Cáncer, Moreto y Matos Fragoso) comparte un score de 51 con *La misma conciencia acusa*¹⁵ (Moreto), con expresiones paralelas como «y las lágrimas que vierto», «de las lágrimas que vierto», «señor que me lleve el diablo», «tomen que me lleve el diablo».

CONCLUSIONES

El editor de textos críticos que busque con este procedimiento fijar atribuciones o hipotetizar sobre filiaciones se sentirá quizá defraudado con los resultados: «Válgame el cielo qué escucho» o «Válgame el cielo qué veo» constituyen ciertamente una cadena de palabras compartida entre textos, pero intuitivamente tendemos a

11. Comparten, como hemos visto, una jornada completa (Sloman, 1958, p. 95).

12. El texto en este corpus de *El parecido en la corte* proviene de la extracción automática de la Biblioteca Virtual Cervantes, pero se trata de una refundición del siglo XVIII (Gómez Sánchez-Ferrer, 2017). Quizá debería haberse eliminado del corpus, pero válida en cualquier caso el experimento.

13. «Al parecer, esta fue la única vez en que Lope aprovechó una de sus comedias en la elaboración de otra» (Tyler, 1963, p. 1).

14. Se debe a las repeticiones, sobre todo, del segundo verso del cantarcillo, «¡Deja las avellanicas, moro, / que yo me las varearé!». Véanse las notas a la edición crítica digital en Vega Carpio (2015).

15. Según Elena Di Pinto y Tania de Miguel (2010) se trata de una versión de *El despertar a quien duerme*, de Lope de Vega. Desafortunadamente el presente corpus no la recoge. La mayoría de las obras de Lope provienen del proyecto ARTELOPE, donde no figura esta obra.

descartarla como pasaje paralelo a no ser que esté incluido en una secuencia más larga. Desde el punto de vista lingüístico es una coocurrencia, que como tal podría extraerse mediante otros procedimientos de Lingüística de corpus.

El procedimiento automático aplicado en este estudio es capaz de localizarlas, pero es un hecho que la mayoría son fórmulas estereotipadas de la práctica teatral, que muestran el entramado de (ciertas) obras dramáticas, en particular en las colaboradas, como hemos visto.

La siguiente lista recoge las 15 reutilizaciones más frecuentes de todo este corpus con el *score* entre paréntesis: «qué es esto» (2577), «ay de mí» (1631), «ay infelice de mí» (927), «arma arma guerra guerra» (717), «válgame el cielo qué es esto» (593), «deme los pies vuestra alteza» (520), «eso cómo puede ser» (320), «deme vuestra majestad» (317), «válgame el cielo qué escucho» (295), «la naturaleza humana» (286), «vuestra majestad señor» (284), «no me puedo detener» (276), «la misma naturaleza» (275), «venid mortales venid» (251), «por muchos años y buenos» (249), «válgame el cielo» (249).

No podemos considerarlos falsos positivos solo porque no representen un papel individualizador directo entre obras. De hecho, en este caso sí nos muestran una tendencia en las colaboradas (a falta de un experimento con un corpus balanceado, exhaustivo y más extenso) a usar automatismos y clichés, como los denomina Vega García-Luengos, quien recurre precisamente, junto con los métodos estilométricos, a cadenas de texto repetidas para la nueva atribución de *La monja alférez*. La explicación a este fenómeno se basa en la necesidad de producir mucho y rápido para la escena teatral áurea, lo que «propiciaba la afloración de automatismos y clichés» (Vega García-Luengos, 2021, p. 116), incluso con una estructura similar octosilábica, ya que «cada autor propende a repetir determinadas cadenas de ocho sílabas, que, en buen número de ocasiones, no comunican contenidos de fondo» (Vega García-Luengos, 2021, p. 116). Vemos que la reflexión bien se puede aplicar a los resultados cuantitativos de la producción textual de las comedias de consuno.

Con un corpus completo y diseñado a medida podríamos recuperar sistemáticamente la reutilización de las colaboradas y clasificarlas. En cualquier caso, el método nos permite un estudio integral y cuantitativo de la reutilización, no solo en corpus teatrales, sino ampliable a fuentes históricas, patristicas, clásicas, etc., que puede servir como camino cuantitativo filológico para la difícil filiación textual de la producción dramática del Siglo de Oro.

BIBLIOGRAFÍA

Alviti, Roberta, *I manoscritti autografi delle commedie del Siglo de Oro scritte in collaborazione. Catalogo e studio*, Firenze, Alinea Editrice, 2006.

ARTELOPE = Oleza, Joan, et al., ARTELOPE, en línea, <http://artelope.uv.es>.

- Baczyńska, Beata, «La reescritura en colaboración: *El príncipe perseguido* de Luis Belmonte, Agustín Moreto y Antonio Martínez de Meneses frente a la comedia fuente *El gran duque de Moscovia* y *Emperador perseguido* de Lope de Vega», *Rilce. Revista de filología hispánica*, 35.1, 2019, pp. 779-805. DOI: <https://doi.org/10.15581/008.35.3.779-805>.
- Büchler, Marco, Philip R. Burns, Martin Müller, Emily Franzini y Greta Franzini, «Towards a Historical Text Re-use Detection», en *Text Mining*, ed. Chris Biemann y Alexander Mehle, Cham, Springer International Publishing, 2014, pp. 221-238. DOI: https://doi.org/10.1007/978-3-319-12655-5_11.
- Büchler, Marco, Annette Geßner, Thomas Eckart y Monica Berti, «Measuring the Influence of a Work by Text Re-use», en *Digital Classicist Supplement: Bulletin of the Institute of Classical Studies*, ed. Stuart Dunn y Simon Mahony, London, University of London, 2013, pp. 63-79.
- Di Pinto, Elena (prólogo y notas) y Miguel Magro, Tania de (aparato crítico), *La misma conciencia acusa*, en *Primera parte de comedias de Agustín Moreto*, vol. II, dir. María Luisa Lobato, coord. Judith Farré Vidal, Kassel, Edition Reichenberger, 2010.
- Escudero Baztán, Juan Manuel, «Reescrituras dramáticas áureas de *La fuerza de la sangre* de Cervantes», *Anales Cervantinos*, 45, 2013, pp. 155-174. DOI: <https://doi.org/10.3989/anacervantinos.2013.007>.
- Fernández Mosquera, Santiago, *Calderón: texto, reescritura, significado y representación*, Madrid / Frankfurt am Main, Iberoamericana / Vervuert, 2015.
- Franzini, Greta, Marco Passarotti, Maria Moritz y Marco Büchler, «Using and Evaluating TRACER for an *Index fontium computatus* of the *Summa contra Gentiles* of Thomas Aquinas», en *Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it 2018: 10-12 December 2018, Torino*, ed. Elena Cabrio, Alessandro Mazzei y Fabio Tamburini, Torino, Accademia University Press, 2019, pp. 199-205.
- Gómez Sánchez-Ferrer, Guillermo, «Las refundiciones de *El parecido* o la comedia moretiana (re)escrita en España al gusto de Europa», *Anagnórisis. Revista de investigación teatral*, 15, 2017, pp. 87-110.
- Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann y Mathieu Bastian, «ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software», *PLOS ONE. Public Library of Science*, 9.6, 2014, p. e98679. DOI: <https://doi.org/10.1371/journal.pone.0098679>.
- Moretti, Franco, «"Operationalizing". Or, the Function of Measurement in Literary Theory», *New Left Review*, 84, 2013, pp. 103-119.
- Navarro Colorado, Borja, «Sistemas de anotación semántica para corpus de español», en *Lingüística de corpus en español*, London, Routledge, 2022, pp. 419-434. DOI: <https://doi.org/10.4324/9780429329296-32>.

- Pichler, Axel, y Nils Reiter, «From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities», *Journal of Cultural Analytics*, 7.4, 2022, s. p. DOI: <https://doi.org/10.22148/001c.57195>.
- Pinillos, Carmen, «Intertextualidad y reescritura paródica en la comedia burlesca: *Céfalo y Pocris* de Calderón», en *Calderón 2000. Homenaje a Kurt Reichenberger en su 80 cumpleaños*, ed. Ignacio Arellano, Kassel, Edition Reichenberger, 2002, pp. 1107-1116.
- Rodríguez Gallego, Fernando, «Aproximación a la reescritura de comedias de Calderón de la Barca», en *Como en la antigua, en la edad nuestra. Presencia de la tradición en la literatura española del Siglo de Oro*, ed. Natalia Fernández Rodríguez, Bellaterra, Universitat Autònoma de Barcelona, 2010, pp. 157-193.
- Ruano de la Haza, José María, *La primera versión de La vida es sueño*, Liverpool, Liverpool University Press, 1992.
- Ruano de la Haza, José María, «Las dos versiones de *El mayor monstruo del mundo*, de Calderón», *Criticón*, 72, 1998 (*Siglo de Oro y reescritura. Tomo I: Teatro*), pp. 35-47.
- Sáez, Adrián J., «Reescritura e intertextualidad en Calderón: *No hay cosa como callar*», *Criticón*, 117, 2013, pp. 159-176. DOI: <https://doi.org/10.4000/criticon.237>.
- Sáez Raposo, Francisco, «El equilibrio imposible del teatro de Agustín Moreto entre el plagio y el canon», en *Como en la antigua, en la edad nuestra. Presencia de la tradición en la literatura española del Siglo de Oro*, ed. Natalia Fernández Rodríguez, Bellaterra, Universitat Autònoma de Barcelona, 2010, pp. 195-226.
- Sánchez Jiménez, Antonio, «Relaciones inter- e intratextuales de un auto de Calderón de la Barca: *La cena del rey Baltasar*», en *Diferentes y escogidas. Homenaje al profesor Luis Iglesias Feijoo*, ed. Santiago Fernández Mosquera, Madrid / Frankfurt am Main / Iberoamericana / Vervuert, 2014, pp. 475-496.
- Sloman, Albert E., *The Dramatic Craftsmanship of Calderón. His Use of Earlier Plays*, Oxford, The Dolphin Book Company, 1958.
- Tyler, Richard W., «Otra mirada a dos comedias de Lope de Vega: *La corona de Hungría* y *Los pleitos de Ingalaterra*», *Hispanófila*, 18, 1963, pp. 21-75.
- Vega Carpio, Lope de, *La dama boba: edición crítica y archivo digital*, Bolonia / Barcelona, PROLOPE / Alma Mater Studiorum Università di Bologna / CRR-MM, 2015. DOI: <https://doi.org/10.6092/UNIBO/LADAMABOBA>.
- Vega García-Luengos, Germán, «La reescritura permanente del teatro español del Siglo de Oro: nuevas evidencias», *Criticón*, 72, 1998 (*Siglo de Oro y reescritura. Tomo I: Teatro*), pp. 11-34.
- Vega García-Luengos, Germán, «Usos y usufructos de Calderón en las comedias colaboradas», en *La comedia escrita en colaboración en el teatro del Siglo de Oro*, ed. Juan Matas Caballero, Valladolid / Olmedo, Ediciones Universidad de Valladolid / Ayuntamiento de Olmedo, 2017, pp. 181-202.

- Vega García-Luengos, Germán, «Juan Ruiz de Alarcón recupera *La monja alférez*», en *Sor Juana Inés de la Cruz y el teatro novohispano*, Rafael González Cañal y Almudena García González, Cuenca, Ediciones de la Universidad de Castilla-La Mancha, 2021, pp. 89-149.
- Vila Carneiro, Zaida, «Reutilización de material literario en las comedias de Calderón: un análisis desde sus piezas más tempranas», *Rilce. Revista de filología hispánica*, 33.2, 2017, pp. 726-747. DOI: <https://doi.org/10.15581/008.33.2.726-47>.
- Vitse, Marc, «Presentación [Siglo de Oro y reescritura. Teatro]», *Criticón*, 72, 1998 (*Siglo de Oro y reescritura. Tomo I: Teatro*), pp. 5-10.