



Las atribuciones del *Lazarillo* bajo la lupa de la informática

José Luis Madrigal
Graduate Center (CUNY)

RESUMEN:

El presente trabajo se adentra en el enigma de la autoría de *Lazarillo* mediante el empleo de dos programas de software diseñados para detectar similitudes estilísticas: JGAAP y R-stylo. En una escogida lista de textos se determina la mayor o menor proximidad con el librito anónimo y qué candidato tendría más probabilidades de ser su autor. Algunos datos reveladores se extraen de las diversas pruebas y análisis realizados.

PALABRAS CLAVE: *Lazarillo*, autoría, estilometría.

ABSTRACT:

With the help of two software programs designed to spot stylistic similarities, JGAAP and R-stylo, this paper confronts the unsolved conundrum of *Lazarillo's* authorship by sorting out the most likely candidates from a chosen pool of sixteenth century texts. Some surprising conclusions are drawn from the various tests and analysis conducted.

KEYWORDS: *Lazarillo*, authorship, stylometry.

Las atribuciones del *Lazarillo* se han disparado en estas últimas décadas. A los tradicionales candidatos propuestos desde el siglo XVII hay que sumar últimamente una pléthora de aspirantes con mayor o menor credibilidad¹. La verdad es que a cada poco amanece con una nueva autoría u ocurrencia². Yo mismo defendí en su momento (y con

1.- Aquí pongo una lista, seguramente incompleta, de algunos de los candidatos propuestos desde principios del siglo XVII: Fr. Juan de Ortega (Sigüenza, 1605; Bataillon, 1954; Guillén, 1966; Alatorre, 2002; Delfín Val, 2009; García Jiménez, 2013); Diego Hurtado de Mendoza (Taxandro, 1607; Schott, 1608; González Palencia, 1944; Agulló, 2010; Corencia Cruz, 2022), Juan de Valdés (M. J. Asensio, 1959; Calvo, 2020), Alfonso de Valdés (Morel Fatio, 1888; Ricapito, 1976; Rosa Navarro, 2002), Sebastián de Horozco (Cejador, 1914; Márquez Villanueva, 1957); Hernán Núñez de Toledo (Rumeau, 1964); Lope de Rueda (Abrams, 1964), Fernando de Rojas (Mancing, 1976; Morcillo Pérez, 2022), Pedro de Rúa (Marasso, 1955), Gonzalo Pérez (Brenes Carrillo, 1986); Torres Naharro (Forcadas, 1994); Juan Maldonado (Colahan y Rodríguez, 1995), Francisco Cervantes de Salazar (Madrigal, 2003), Juan Luis Vives (Calero, 2005), Juan Arce de Otálora (Madrigal, 2008), Francisco de Encinas (Labarre, 2006; Rodríguez López-Vázquez y Rodríguez López-Abadía, 2015), Juan Bernal Díaz de Lucio (Morales, 2022). Consenso no parece haber en el asunto ni tampoco mucha cordura en algún caso.

2.- Un estado de la cuestión muy completo sobre las atribuciones del *Lazarillo* se puede leer en ROSA y SUÁREZ (2016). Este trabajo presenta también un excelente y exhaustivo análisis de estilística computacional en relación con muchos de los candidatos a la autoría del *Lazarillo*.

una vehemencia de la que ahora me arrepiento) la candidatura de Francisco Cervantes de Salazar y en la actualidad sugiero de manera más ponderada (aunque con no menos convencimiento) la del licenciado Juan Arce de Otálora. No hablaré por otros. Si yo me caí del burro con respecto al humanista toledano y me subí al carro del jurista vallisoletano no fue por una genial intuición, sino por la fuerza incontestable de los datos internos. A falta de un testimonio en forma de carta, testamento o cualquier otro documento legal, solo nos queda el texto mismo. Aulio Gelio, en el siglo II, ya dejó dicho que la mejor manera de confirmar la autoría de las comedias de Plauto era mediante el análisis estilístico³. Un lector competente distingue de inmediato un soneto de Garcilaso y de Góngora, una novela escrita por Galdós o por Valera, la prosa de Valle Inclán de la de Baroja. Otra cosa es demostrar empíricamente quién escribe qué si nos falta la firma del autor o la obra aparece o se publica anónima. En cuanto algo así sucede, hasta la candidatura más inequívoca se puede rebatir. La historia de la literatura no carece de ejemplos a este respecto. Para Isidoro Bosarte, Martín Fernández de Navarrete o Bartolomé José Gallardo la autoría cervantina de *La tía fingida* encontrada en el manuscrito Porras era evidentísima, pero ya sabemos la suerte que corrió la desenfadada novelita a partir de las dudas presentadas por Foulche-Delbosc. Inútil fue enarbolar en años posteriores docenas de paralelismos o guiarse por el sentido común. Desde entonces la obra ha permanecido anónima y solo recientemente, y con reparos, alguno se atreve a publicarla atribuida a Cervantes⁴. ¿Hay algún modo de paliar la duda metódica? Pues creo que el único remedio que tenemos para ello es la estadística. Los avances son cada vez mayores en este terreno. Hoy en día disponemos de dos programas de estilística computacional al alcance de cualquiera: JGAAP (Java Graphical Authorship Attribution Program) desarrollado por Patrick Juola en la Universidad de Duquesne⁵ y STYLO (Stylometry with R), creado por Maciej Eder y su equipo del Instituto de Lengua Polaca (Institute of Polish Language)⁶. Uno y otro programa están disponibles en la red desde hace años y son de empleo muy sencillo. JGAAP ofrece al usuario un amplio abanico de métodos y rasgos lingüísticos cuantificables, mientras que STYLO analiza la frecuencia de palabras y n-gramas de caracteres mediante métricas de distancia de análisis estadístico multivariante, pero con la gran ventaja de visualizar los resultados gracias a una muy atractiva interfaz gráfica.

1. Una primera prueba: «*La tía fingida*»

No entraré en ningún detalle técnico⁷. Pongámonos manos a la obra y veamos qué pasa con *La tía fingida* a la luz de estos dos programas informáticos. En Apéndice I incluyo el corpus de textos y autores. Empezaré la prueba con JGAAP. El método elegido es *K-Nea-*

3.- Según Aulio Gelio, el polígrafo Marco Varrón había atribuido varias comedias a Plauto «por la índole del estilo y el carácter de los chistes» (Noctes Atticae, III, 3).

4.- Yo eché mi cuarto a espadas en defensa de la autoría cervantina con una batería de paralelismos y colocaciones léxicas abrumadora, o eso me parecía (MADRIGAL, 2002), pero los escépticos siguen ahí. Véase, por ejemplo, BARAS ESCOLÁ.

5.- <<http://evllabs.github.io/JGAAP/>>

6.- <<http://maciejeder.org/projects/stylo/>>

7.- Para quien quiera más información puede consultar JUOLA (2008) y EDER (2016, 2020). Un buen resumen sobre estilística computacional y atribución textual en SAVOY. El artículo de ROSA centrado en las atribuciones del *Lazarillo* es también muy útil.

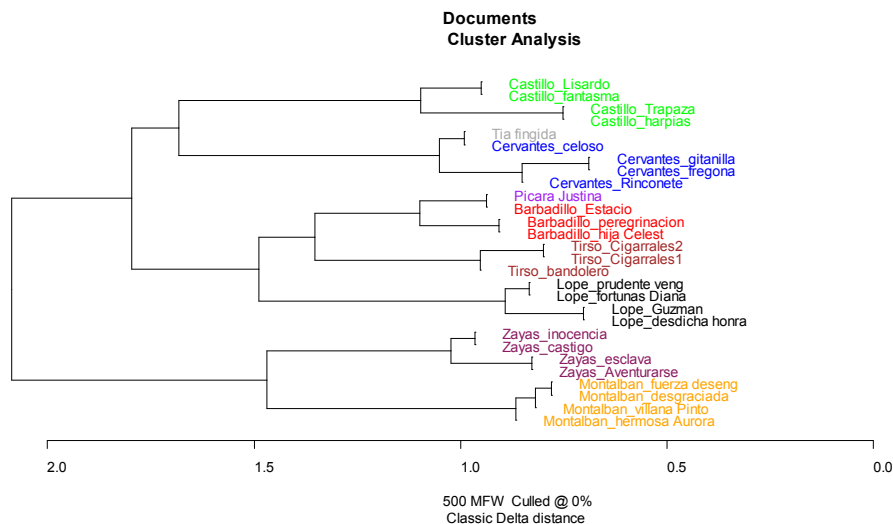
rest Neighbor Driver with metric Cosine Distance, algo así como Vecinos más Próximos con Distancia Métrica de Coseno, aunque a partir de ahora emplearé la abreviatura KNN. Dos son los rasgos analizados: n-gramas de 4 caracteres (C4) y n-gramas de una palabra (W1). El programa solo criba aquí los cinco primeros textos, siendo 1.5 el más próximo a *La tía fingida*.

4 n-gramas de caracteres C4	1 n-grama de palabras W1
1. SALAS BARBADILLO <i>La hija de la Celestina</i> 1.0312 2. <i>Pícara Justina</i> 1.0625 3. CERVANTES <i>La Ilustre fregona</i> 1.125 4. CERVANTES <i>El celoso extremeño</i> 1.25 5. CERVANTES <i>Coloquio de los perros</i> 1.5	1. CERVANTES <i>La Ilustre fregona</i> 1.03125 2. CERVANTES <i>Rinconete</i> 1.0625 3. CERVANTES <i>La gitanilla</i> 1.125 4. CERVANTES <i>Coloquio de los perros</i> 1.25 5. CERVANTES <i>El celoso extremeño</i> 1.5

De los 25 textos utilizados en el experimento las cinco novelas de Cervantes se llevan la palma en las dos categorías. Nótese, además, que en el cotejo con las palabras más frecuentes (W1), las *Novelas ejemplares* copan todos los puestos. Si en lugar de la distancia de coseno utilizamos la distancia Hellinger, el resultado es muy parecido:

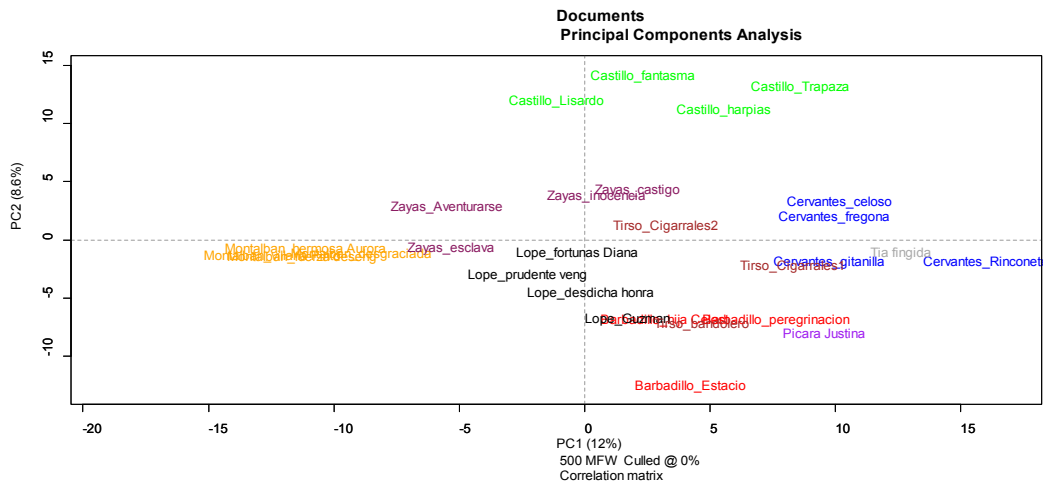
C4	W1
1. SALAS BARBADILLO <i>El sagaz Estacio</i> 1.03 2. SALAS BARBADILLO <i>La hija de la Celest.</i> 1.06 3. CERVANTES <i>El celoso extremeño</i> 1.125 4. CERVANTES <i>La gitanilla</i> 1.25 5. CERVANTES <i>Coloquio de los perros</i> 1.5	1. CERVANTES <i>La Ilustre fregona</i> 1.03125 2. ZAYAS <i>La esclava de su amante</i> 1.0625 3. CERVANTES <i>Coloquio de los perros</i> 1.125 4. CERVANTES <i>El celoso extremeño</i> 1.25 5. CERVANTES <i>La gitanilla</i> 1.5

Vayamos ahora al programa STYLO. Escojo en un primer experimento el análisis de conglomerados (Cluster Analysis) con métrica de distancia Delta propuesto por Burrows:

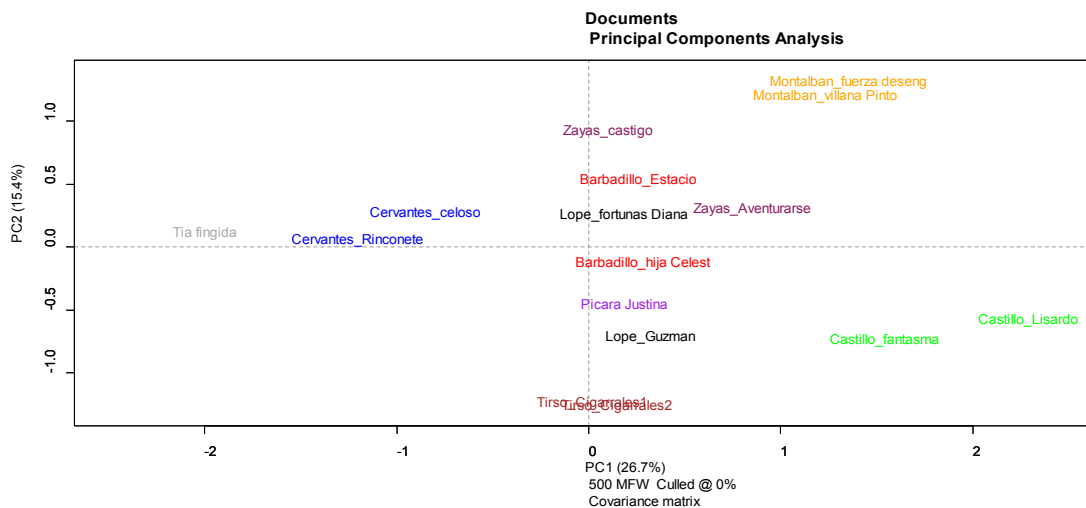


La clasificación que se hace de los textos es de lo más exacta. En cada horquilla se agrupan los textos de acuerdo con su autor. *La tía fingida* está emparejada con *El celoso extre-*

meño y dentro de la misma rama que las otras novelas cervantinas. Resulta curioso que *La pícaro Justina*, la otra obra de autoría disputada, aparezca asociada con las novelas de Salas Barbadillo. En este caso hemos empleado las 500 palabras más frecuentes. Veamos ahora qué pasa si cambiamos de método y queremos visualizar mejor la distancia entre textos. El análisis de componentes principales (PCA) es el más adecuado en este caso. STYLO ofrece varias matrices. Empecemos con la matriz de correlación.

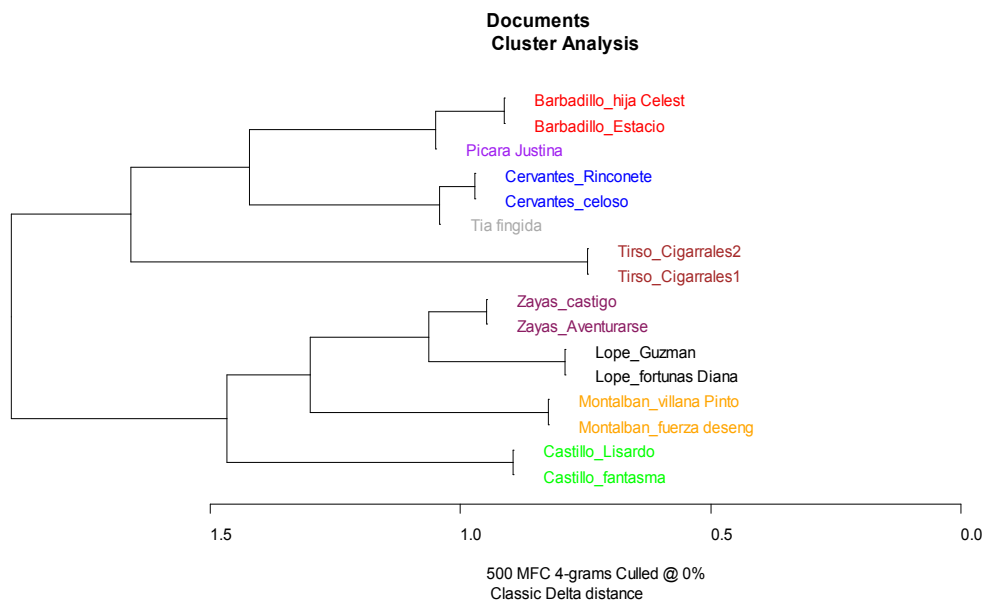


En el diagrama vemos que *La tía fingida* (en gris) está pegada a *Rinconete* y se sitúa cerca de las otras novelas de Cervantes aquí analizadas. No se observa ninguna discrepancia notable, salvo quizá en los textos de Tirso. A fin de ver todo esto con más claridad, limpiemos el corpus de algunos textos redundantes y, en lugar de la matriz de correlación, empleemos la matriz de covarianza:



Esta vez los dos textos de Tirso están superpuestos y en general la proximidad de cada texto analizado está en consonancia con quién lo escribió, a excepción quizá de las dos novelas de Lope. En cuanto a *La tía fingida*, el diagrama confirma la estrecha proximidad

que tiene con las otras dos *Novelas ejemplares* que se encontraron en el manuscrito Porras. Un último análisis de conglomerados, esta vez con n-gramas de caracteres, clasifica con precisión todos los textos:



El experimento nos ha demostrado que tanto JGAAP como STYLO son muy fiables en la discriminación textual. No es seguramente la panacea universal para resolver todo problema de autoría, pero sí parece claro que determina con bastante exactitud la semejanza entre textos. Pasemos, sin más dilación, al *Lazarillo*.

2. Análisis con textos de 10.000 palabras

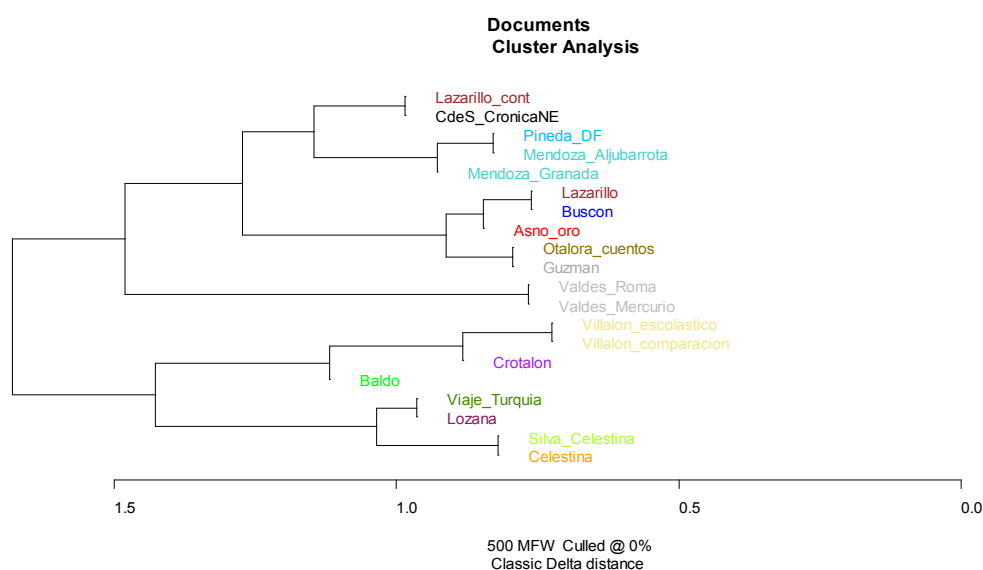
El texto publicado en 1554 es único en su género, por lo menos en el momento de su aparición. En todo caso, hagamos la misma prueba que hemos hecho con *La tía fingida*. [Consúltense en Apéndice I el corpus empleado]. La mayoría de los textos están escritos en el siglo XVI y tienen unas 10,000 palabras. Empecemos una vez más con JGAAP. El método vuelve a ser KNN con distancia coseno:

C4	W1
1. PINEDA <i>Diálogos familiares</i> 1.03125	1. <i>Diálogo de las transformaciones</i> 1.03125
2. <i>Diálogo de las transformaciones</i> 1.0625	2. Narváez <i>Diálogo del capón</i> 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. OTALORA Cuentos 1.125
4. OTALORA Cuentos 1.25	4. <i>Asno de oro</i> 1.25
5. <i>Asno de oro</i> 1.5	5. <i>Segunda Parte Lazarillo</i> 1.5

Otálora obtiene un segundo puesto en las secuencias de 4 n-gramas y un tercero en las palabras más frecuentes. Si probamos con la distancia Hellinger, los resultados son aún mejores:

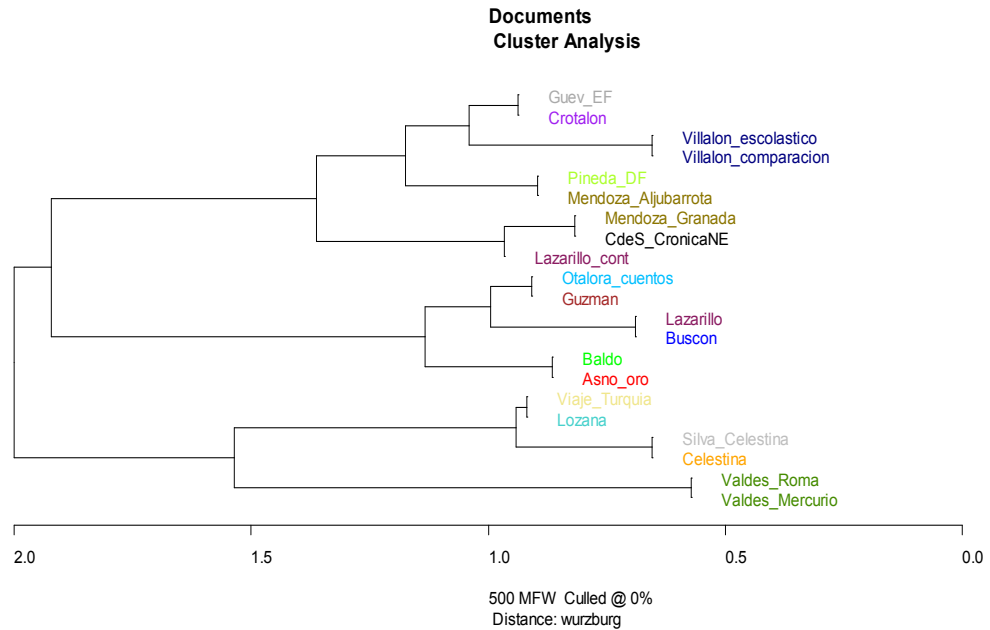
C4	W1
1. <i>Viaje Turquía</i> 1.03125	1. VALDES <i>Mercurio</i> 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. <i>Segunda parte Lazarillo</i> 1.125	3. <i>Segunda parte Lazarillo</i> 1.125
4. OTALORA Cuentos 1.25	4. OTALORA Cuentos 1.25
5. <i>Asno de oro</i> 1.5	5. <i>Asno de oro</i> 1.5

No está nada mal llevarse tres «platas» y un «bronce» cuando se compite con *El Asno de oro*, uno de los modelos en que se basa el *Lazarillo*, y *La Segunda Parte*, que es, al fin y al cabo, su continuación. Aclaro que el texto de Otálora lo integran cuatro cuentos extraídos de *Coloquios de Palatino* y *Pinciano*. Más adelante daré mayor información. Vayamos ahora a STYLO. Lo inicio con un análisis de conglomerados y las 500 palabras más frecuentes:

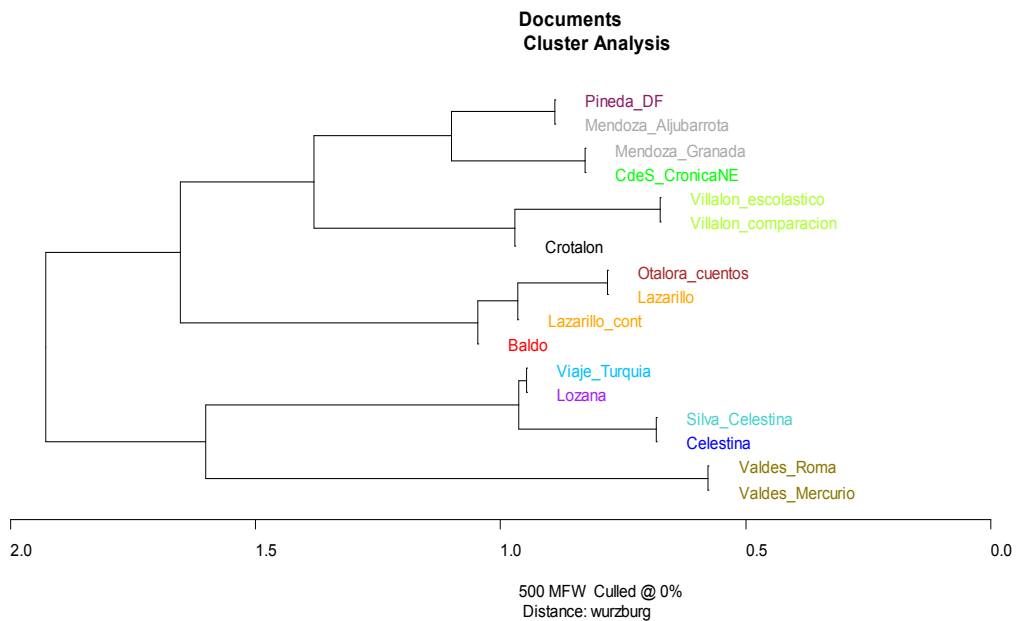


El *Lazarillo* se empareja con el *Buscón* y entronca con el *Asno de oro*, el *Guzmán de Alfarache* y los cuentos de Otálora. *La Segunda Parte* se sitúa junto a la *Crónica* de Cervantes de Salazar y entronca con los *Diálogos familiares* de Pineda, el *Sermón de Aljubarrota* y la *Guerra de Granada*. Los diálogos de *Mercurio* y *Caronte* y las *Cosas acaecidas en Roma* de Alfonso de Valdés están en una misma horquilla, lo mismo que las dos obras de Villalón y el *Crótalon*.

Probemos ahora con la distancia Delta de coseno Wurzburg, en lugar de la tradicional de Burrows:

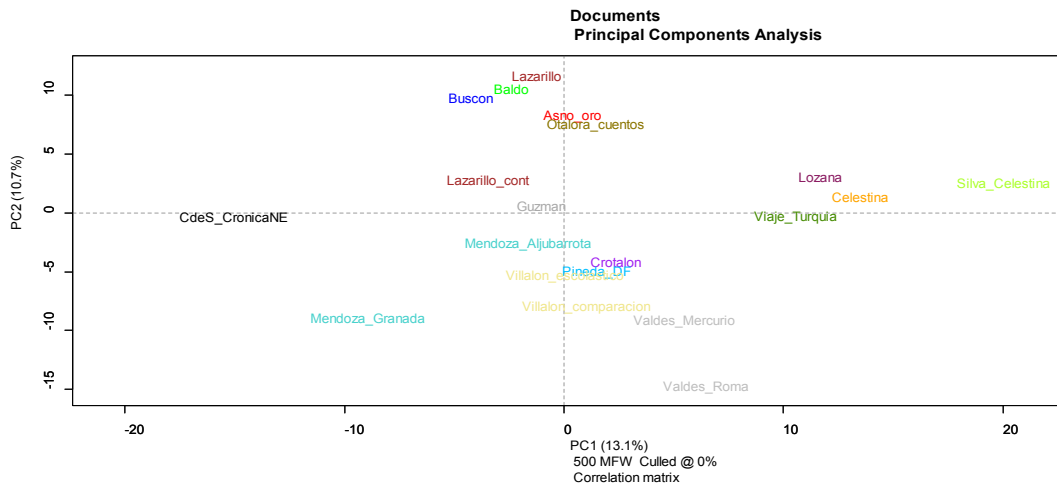


Los cuentos de Otálora permanecen una vez más en la misma horquilla de *Lazarillo* junto a las novelas de género picaresco. Si eliminamos del corpus el *Buscón* y el *Guzmán*, textos escritos ya décadas después, así como el *Asno de oro*, traducción hecha al menos treinta años antes de la publicación del librito anónimo, los cuentos de Otálora y el *Lazarillo* quedan entonces emparejados entre sí:

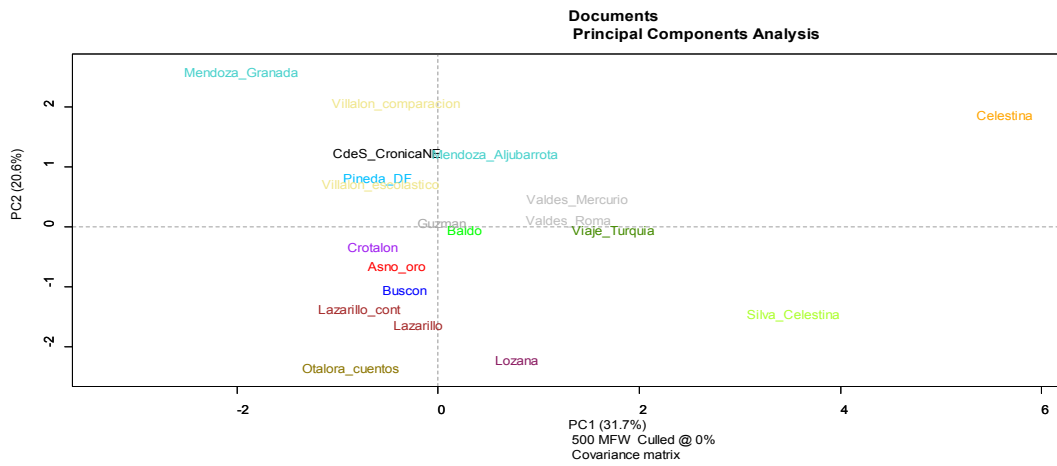


Visualicemos ahora las distancias entre textos con varios análisis de PCA (Análisis de componentes principales):

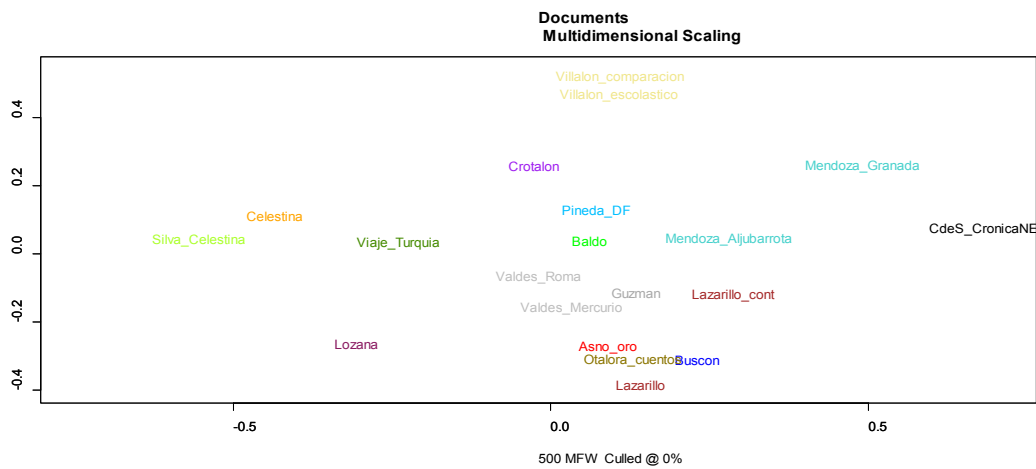
1) PCA Matriz de correlación (500MFW)



2) PCA Matriz de covarianza (500MFW)



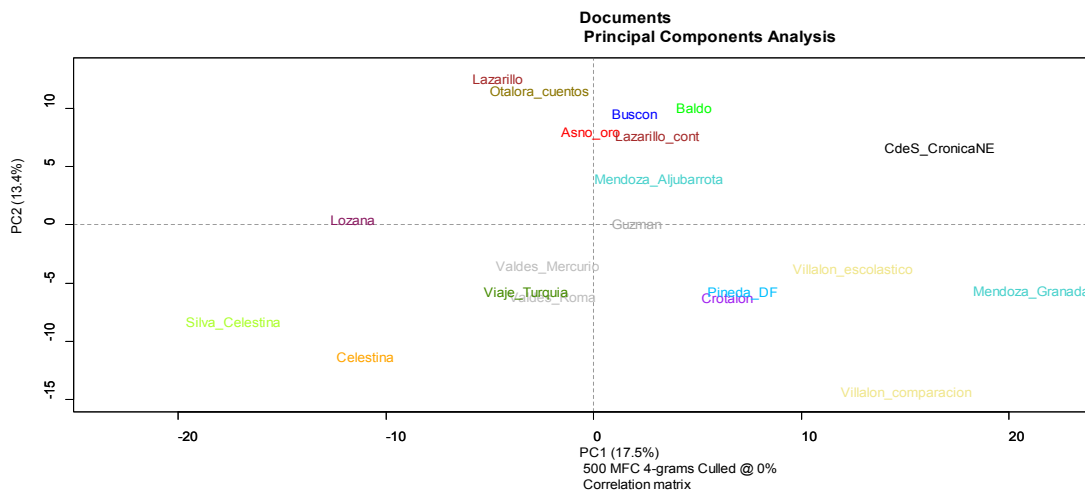
3) PCA Escalamiento multidimensional (MDA) 500 MFW



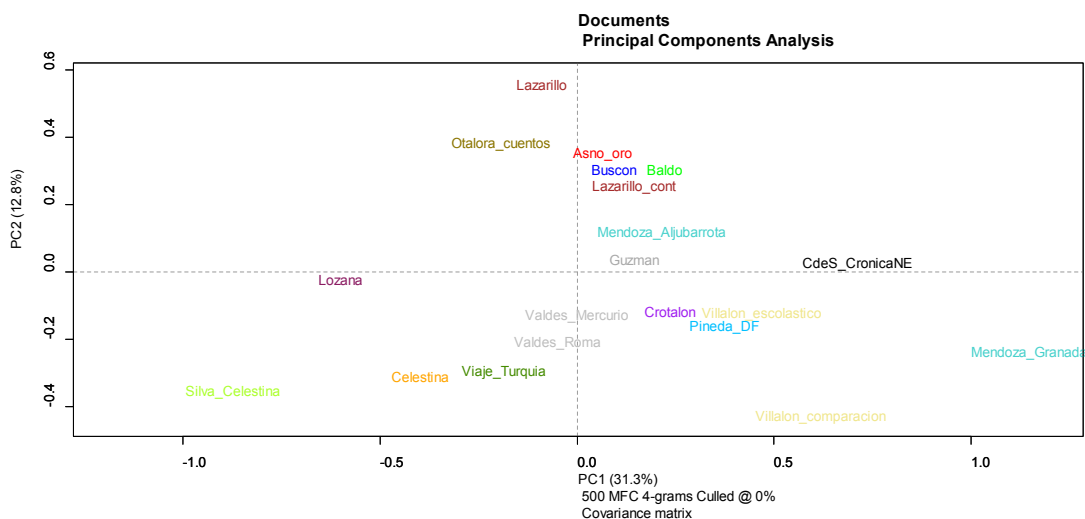
En los tres gráficos el texto del *Lazarillo* y los cuentos de Otálora están muy próximos. En cuanto a las novelas que se enmarcan dentro del modelo picaresco, caso del *Buscón* o del *Asno de oro*, la cercanía es también notable. El *Guzmán de Alfarache* orbita cerca, pero no tanto como se supondría. Es de notar que la continuación del *Lazarillo* solo está próxima en el segundo diagrama. Obsérvese asimismo que en los tres diagramas existe una distancia considerable entre el librito anónimo y los textos de Valdés, Mendoza y Cervantes de Salazar.

Hagamos la misma operación, pero esta vez con n-gramas de caracteres:

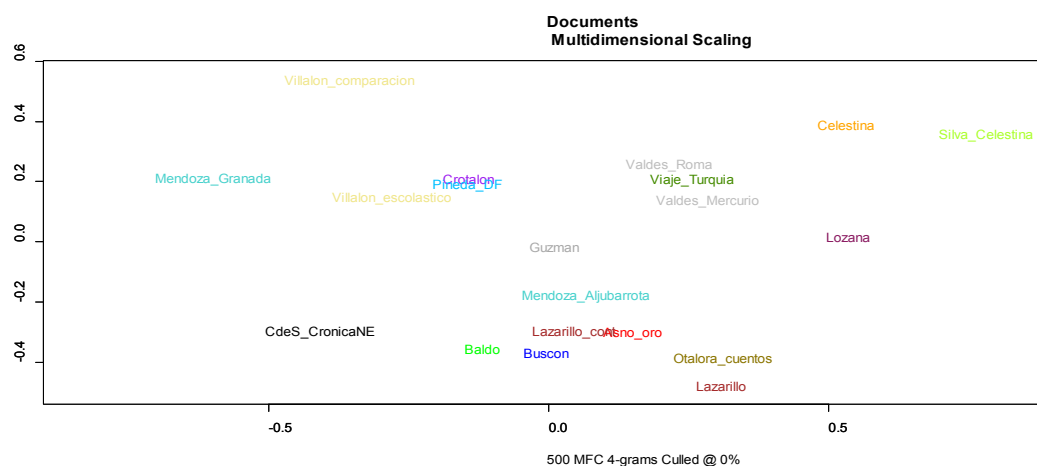
1) PCA Matriz de correlación (500MFC 4-grams)



2) PCA Matriz de covarianza (500MFC 4-grams)



3) PCA Escalamiento multidimensional (500 MFC 4-grams)

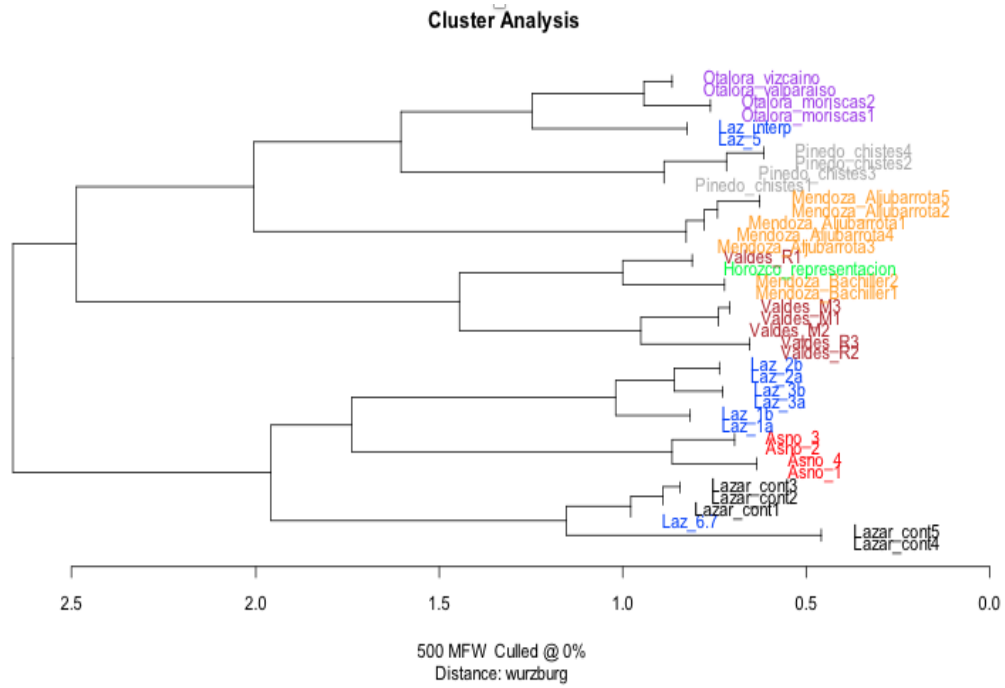


En los tres diagramas el *Lazarillo* y los cuentos de Otálora son los más próximos entre sí. Su grado de cercanía es ciertamente asombroso. Ni siquiera las obras picarescas superan en ningún caso el texto de Otálora. Los otros candidatos propuestos (Valdés, Mendoza y Cervantes de Salazar) quedan otra vez muy lejos.

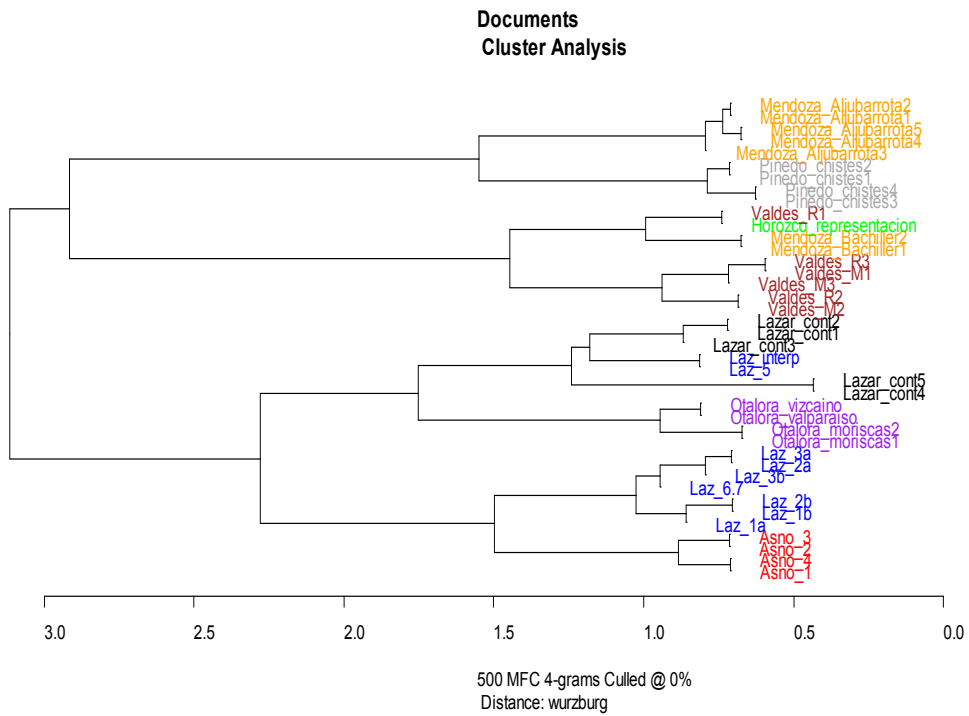
3. Análisis con textos de 2.000 palabras

Hasta aquí hemos examinado textos de 10.000 palabras. Dado que el *Lazarillo* tiene unas 18.000, dividámoslo ahora en unidades de 2.000 palabras. Así, los tres primeros tratados tendrán dos textos cada uno (Laz_1a y Laz_1b, Laz_2a y Laz_2b, Laz_3a y Laz_3b), mientras que el tratado quinto (el del buldero) formará una sola unidad (Laz_5). Dejaré fuera el cuarto tratado y crearé un solo texto con el tratado VI y VII (Laz_6 7). Advierto que este último ((Laz_6 7)) tiene escasamente 1000 palabras. Me to también en el corpus las Interpolaciones (Laz_interp) de la edición de Alcalá, pese a contar solo con 1500 palabras. Haré la misma división con el resto de textos del corpus. Por ejemplo, los cuentos de Otálora sacados de *Coloquios* son cuatro: el cuento del corredor (Otalora_corredor), el del vizcaíno (Otalora_vizcaino), el cuento del hidalgo burlado (Otalora_valparaiso) y el cuento de las hermanas moriscas y los dos estudiantes, que irá dividido en dos unidades (Otalora_moriscas1 y Otalora_moriscas2). Cada texto cuenta con 2000 palabras aproximadamente. Elimino del corpus el *Guzmán de Alfarache* y el *Buscón*, además de todas aquellas obras que se vieron sistemáticamente fuera de la órbita del *Lazarillo* en las pruebas anteriores. Quedan excluidos así la *Celestina* original, la *Segunda Celestina* de Feliciano de Silva, las obras de Villalón, la *Crónica* de Cervantes de Salazar, *Diálogos familiares* de Pineda, *Viaje de Turquía*, *La lozana andaluza* y *La guerra de Granada* de Mendoza. Conservo los diálogos de Valdés por el mucho ruido que ha habido en torno a su candidatura, aunque deberían también excluirse. Otras novedades en este corpus son el *Libro de chistes* de Luis de Pinedo y la *Representación evangélica de San Juan* de Sebastián de Horozco. En Apéndice I puede verse el corpus completo. Empiezo con el tradicional análisis de conglomerados:

500 MFW (n-gramas de palabras)

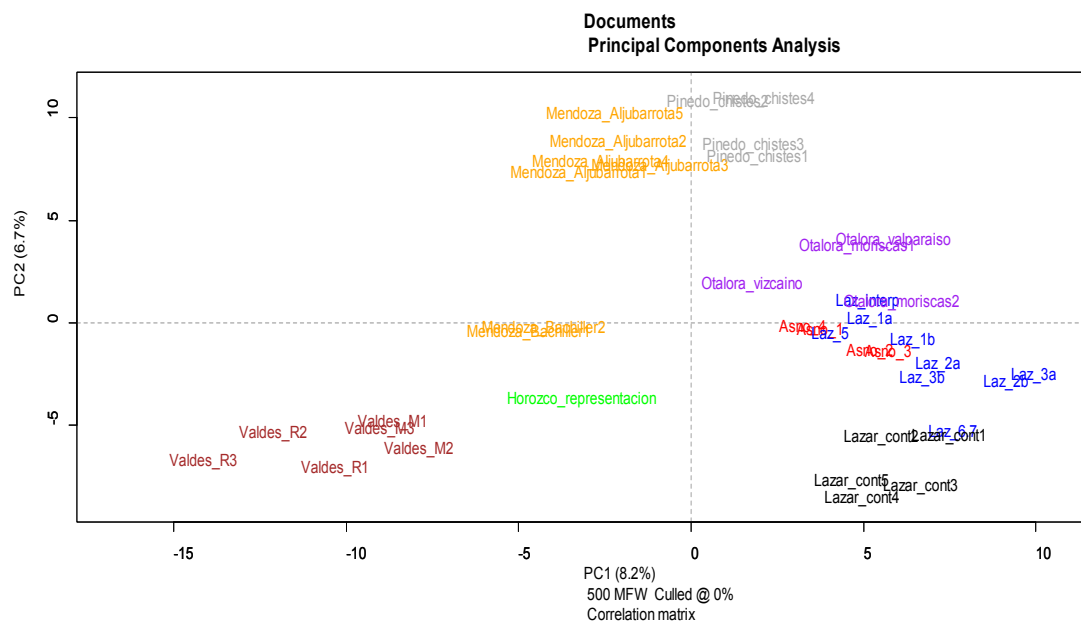


500 MFC 4 (n-gramas de 4 caracteres)

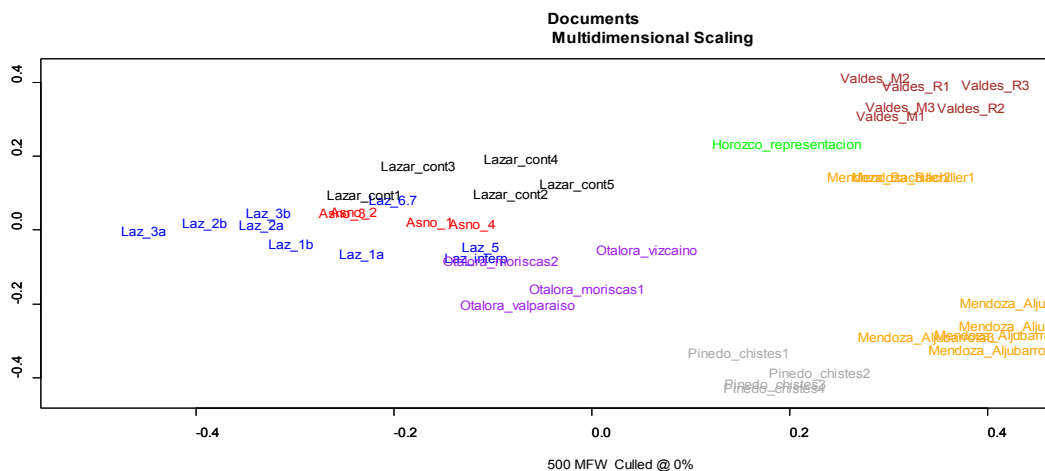


Se dan varias peculiaridades importantes en estos dos primeros dendrogramas. Por lo pronto, en el primero (500 palabras más frecuentes) los cuentos de Otálora están en la misma horquilla del tratado del buldero (Laz_5) y las interpolaciones de Alcalá (Laz_interp), mientras que en el segundo (n-gramas de 4C), el tratado V y las interpolaciones vuelven a desgajarse del resto de tratados, aunque esta vez incrustados en los textos de la *Segunda Parte*. Asimismo, Otálora en el segundo análisis queda agrupado junto con *Lazarillo*, *Asno de oro* y *Segunda parte*. El resto de textos aparecen desligados tanto de *Lazarillo* como de su continuación en ambos dendrogramas. Pasemos a los análisis de componentes principales con el fin de visualizar mejor la distancia entre los distintos textos:

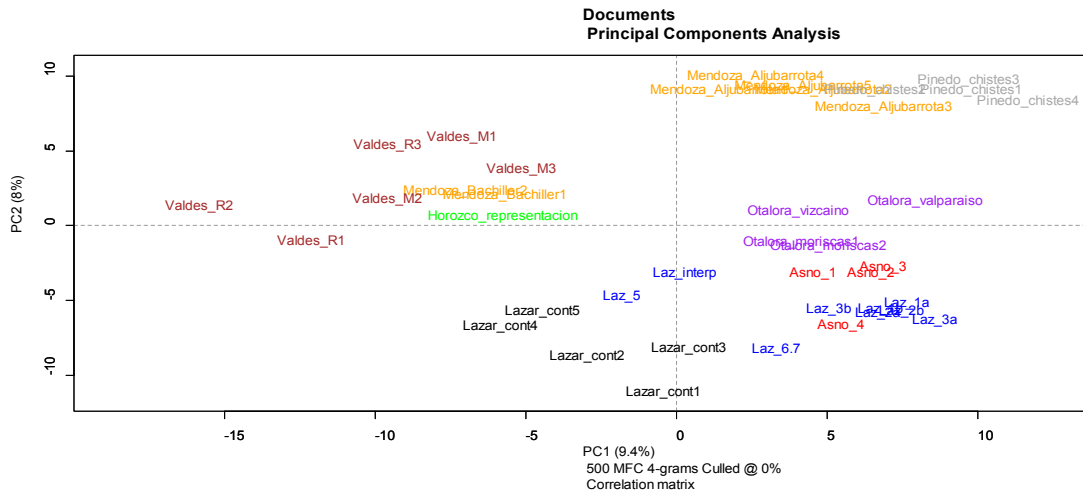
1. PCA 500 MFW - Matriz de correlación (frecuencia de palabras)



2. PCA 500 MFW Escalamiento Multidimensional (MDS)



3. PCA (500 MFC 4-grams). Matriz de correlación.



Hay varias constantes en los tres diagramas. Así, los textos de Mendoza, Valdés, Pinedo y el de Horozco aparecen siempre muy distanciados de los dos *Lazarillos*. Es de notar también la distancia observada entre los textos de *Bachiller Arcadia* y los del *Sermón de Aljubarrota*. En contraposición, *Asno de oro*, Otálora, *Lazarillo* y *La Segunda parte* se agrupan en un mismo espacio, aunque hay diferencias de grado. Así, todos los textos de *La Segunda parte* está casi siempre a cierta distancia, mientras que los del *Asno de oro* y Otálora aparecen en general más próximos a los textos del *Lazarillo*. En el primer gráfico Otalora_moriscas2 está superpuesto a las interpolaciones (Laz_interp) y justo encima del primer tratado (Laz_1a) y en el segundo observamos el mismo fenómeno, pero esta vez entre Otalora_moriscas2, Laz_5 y Laz_interp. Los análisis con n-gramas de cuatro palabras ofrecen resultados no muy distintos (v. Apéndice III).

5. Hurtado de Mendoza, Luis de Pinedo y Arce de Otálora

A tenor de lo visto, ni don Diego Hurtado de Mendoza (al menos el don Diego a quien se le atribuyen la *Carta del Bachiller Arcadia* y *El sermón de Aljubarrota glosado*) ni Alfonso de Valdés residen, por así decir, en el mismo barrio del autor del *Lazarillo*. Lo mismo puede decirse de Luis de Pinedo. Ciertamente su *Libro de chistes* tiene varias referencias explícitas al *Lazarillo*, pero sin descartar la pertenencia a un mismo círculo, no parece haber otra relación con el librito anónimo⁸. En un rápido examen realizado con el programa JGAAP se ve que *Asno de oro*, *La segunda parte del Lazarillo* y los cuentos de Otálora copan consistentemente las primeras posiciones y ni una sola vez asoma Pinedo, mientras que Mendoza y Alfonso de Valdés aparecen de manera esporádica. En Apéndice II van todos los resultados; debajo solo incluyo los más pertinentes:

8.- Considérese, por lo demás, que el libro parece ser una recopilación hecha no solo por Pinedo, sino por otros, tal como se infiere del título en latín: *Liber facietarum et similitudinum Ludovici de Pinedo et amicorum*.

1) Lazarillo (10.000 palabras)

C4	W
1. MENDOZA <i>Sermón de Aljubarrota</i> 1.03125	1. VALDES (<i>Mercurio</i>) 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. <i>Segunda Parte Lazarillo</i> 1.125
4. OTALORA Cuentos 1.25	4. OTALORA Cuentos 1.25
5. <i>Asno de oro</i> 1.5	5. <i>Asno de oro</i> 1.5

2) PRIMER TRATADO

Laz_1a (2.000 palabras)

C4	W
1. MENDOZA <i>Sermón de Aljubarrota</i> 1.03125	1. <i>Lozana andaluza</i> 1.03125
2. <i>Segunda Parte Lazarillo</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. CERVANTES <i>Galatea</i> 1.125	3. <i>Segunda Parte Lazarillo</i> 1.125
4. OTALORA Cuentos 1.25	4. OTALORA Cuentos 1.25
5. <i>Asno de oro</i> 1.5	5. <i>Asno de oro</i> 1.5

Laz_1b (2.000 palabras)

C4	W
1. <i>Diálogo de las transformaciones</i> 1.03125	1. VALDES (<i>Mercurio</i>) 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. <i>Asno de oro</i> 1.125
4. OTALORA Cuentos 1.25	4. <i>Segunda Parte Lazarillo</i> 1.25
5. <i>Asno de oro</i> 1.5	5. OTALORA Cuentos 1.5

3) TRATADO V (2.000 palabras)

Laz_5

C4	W
1. MENDOZA <i>Sermón de Aljubarrota</i> 1.03125	1. MENDOZA <i>Sermón de Aljubarrota</i> 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. VALDES (<i>Mercurio</i>) 1.0625
3. <i>Asno de oro</i> 1.125	3. <i>Asno de oro</i> 1.125
4. <i>Segunda Parte Lazarillo</i> 1.25	4. <i>Segunda Parte Lazarillo</i> 1.25
5. OTALORA Cuentos 1.5	5. OTALORA Cuentos 1.5

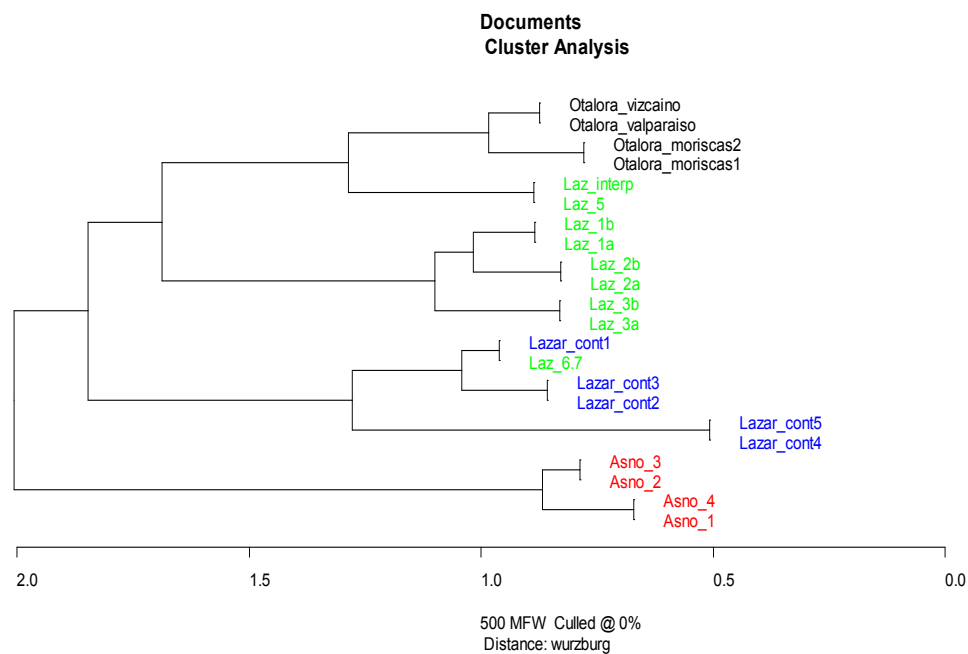
4) INTERPOLACIONES (1530 palabras)

Laz_interp

4 n-gramas de caracteres C4	2 n-gramas de palabras W
1. CERVANTES <i>Galatea</i> 1.03125	1. C DE SALAZAR CNE 1.03125
2. C DE SALAZAR CNE 1.0625	2. VALDES (<i>Mercurio</i>) 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. <i>Asno de oro</i> 1.125
4. <i>Asno de oro</i> 1.25	4. <i>Segunda Parte Lazarillo</i> 1.25
5. OTALORA Cuentos 1.5	5. OTALORA Cuentos 1.5

Los resultados con JGAAP son claros. Vayamos otra vez a STYLO y dejemos fuera tanto el *Libro de chistes* de Pinedo como los textos de Mendoza. Con solos *Asno de oro*, los

dos *Lazarillos* y los cuentos de Otálora, este es el dendrograma que obtenemos con las 500 palabras más frecuentes:



6. Una apostilla final y conclusiones

El objetivo último de cualquier estudio de atribución basado en datos internos, sea el más tradicional de recopilación de paralelismos o bien aquel en donde se calcula la frecuencia de tal o cual rasgo lingüístico mediante algún algoritmo informático, es determinar la mayor o menor similitud entre un conjunto de textos. En principio, no identifica, sino que discrimina y clasifica. Desde hace años defino el idiolecto (o, más bien, el repertorio verbal de un individuo) con tres adjetivos: restringido, singular y recurrente. En efecto, el repertorio verbal de un hablante, por amplio que sea, es siempre finito y, por lo mismo, distinto al de cualquier otro. Es también recurrente, lo cual tampoco puede extrañar, ya que la recurrencia resulta consustancial a la conducta de todo ser vivo. Existe variación, sí, pero ante todo hay repetición en casi todos nuestros actos. Lo que hacemos una vez lo solemos hacer otras muchas veces a lo largo de nuestro curso y discurso vital. La ropa que llevamos puesta, los autores que citamos o los lugares en los cuales alguna vez estuvimos determinan de algún modo nuestra identidad. En 2007 escribí en el buscador de CORDE (el corpus diacrónico de la RAE) esta lista de autores: <Tulio, Plinio, Galeno, Macías, Ovidio, Santo Tomás> y apreté el botón de búsqueda. El resultado me hizo gracia:

Consulta: *Julio y Plinio y Galeno y Macías y Ovidio y San*to Tomás , en 1400-1600, en todos los medios, en CORDE*
 Resultado: 223 casos en 2 documentos.

OBTENCIÓN DE EJEMPLOS Documentos.
 Clasificación: Selección
 Agrupación: Marcas:

Cómo citar el CORPUS Documentos.
 Pantalla: 1 de 1.

	Casos	Año	Autor	Obra	País	Tema	Publicación
<input type="checkbox"/>	217	c 1550	Arce de Otárola, Juan de	Coloquios de Palatino y Pinciano	ESPAÑA	12.Relato extenso diálogo miscelánea y	José Luis Ocasar Ariza, Turner (Madrid), 1995
<input type="checkbox"/>	6	a 1554	Anónimo	Lazarillo de Tormes	ESPAÑA	12.Relato extenso novela y otras formas similares	Francisco Rico, Cátedra (Madrid), 1994

[Ir arriba](#) Pantalla: 1 de 1.

Volví a hacer lo mismo con dos lugares poco habituales: <Tejares y Costanilla>. Esta vez no puse restricciones y espigué del primero al último texto en el corpus de CORDE:

Consulta: *Tejares y Costanilla, en todos los medios, en CORDE*
 Resultado: 5 casos en 2 documentos.

OBTENCIÓN DE EJEMPLOS Documentos.
 Clasificación: Selección
 Agrupación: Marcas:

Cómo citar el CORPUS Documentos.
 Pantalla: 1 de 1.

	Casos	Año	Autor	Obra	País	Tema	Publicación
<input type="checkbox"/>	3	c 1550	Arce de Otárola, Juan de	Coloquios de Palatino y Pinciano	ESPAÑA	12.Relato extenso diálogo miscelánea y	José Luis Ocasar Ariza, Turner (Madrid), 1995
<input type="checkbox"/>	2	a 1554	Anónimo	Lazarillo de Tormes	ESPAÑA	12.Relato extenso novela y otras formas similares	Francisco Rico, Cátedra (Madrid), 1994

[Ir arriba](#) Pantalla: 1 de 1.

Este segundo resultado me aturdió algo más. Recuerdo que por aquel entonces mi candidato era un tal Francisco Cervantes de Salazar. Elegí entonces secuencias en torno a estos nombres y lugares. Varios paralelismos exclusivos me desconcertaron del todo:

1) <... (a) este propósito dice Tulio>

Consulta:	este propósito dice Tulio, en todos los medios, en CORDE
Resultado:	2 casos en 2 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Clasificación:
 Agrupación: Marcas:

Cómo citar el CORPUS **Concordancias.**
 Pantalla: 1 de 1. [Ver párrafos](#)

no es para ir allá. A este propósito dice Tulio en el libro tercero de orat ** c 1550 AÑO Arce de d
 ue, se las alaben. Y a este propósito dice Tulio: "La honra cria las artes". ** a 1554 AUTOR Anónimo

[Ir arriba](#) Pantalla: 1 de 1. [Ver párrafos](#)

Nueva consulta: [CREA](#) [CORDE](#) [Nómina de autores y obras](#) [Ayuda](#).

2) <... dice Plinio... no hay libro... que no tenga alg*>

Consulta:	dice Plinio dist/10 hay libro, en 1400-1600, en todos los medios, en CORDE
Resultado:	2 casos en 2 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Clasificación:
 Agrupación: Marcas:

Cómo citar el CORPUS **Concordancias.**
 Pantalla: 1 de 1. [Ver párrafos](#)

ilces de oír, que como dice Plinio el Mozo, "no hay libro tan malo que no te ** c 1550 AÑO Arce de O:
 de. Y a este propósito dice Plinio que "no hay libro por malo que sea, que ** a 1554 AUTOR Anónimo

[Ir arriba](#) Pantalla: 1 de 1. [Ver párrafos](#)

3) <qué Macías está hecho /estaba ... hecho un Macías>

Consulta:	est* dist/6 Macías dist/3 hecho, en todos los medios, en CORDE
Resultado:	2 casos en 2 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Clasificación:
 Agrupación: Marcas:

Cómo citar el CORPUS **Concordancias.**
 Pantalla: 1 de 1. [Ver párrafos](#)

s qué Macías está hecho el señor mi compañero, si podría entrar si ** c 1550 AÑO Arce de d
 como digo, él estaba entre ellas, hecho un Macías, diciéndoles más ** a 1554 AUTOR Anónimo

4) <... que Ovidio escribió>
(RAE)

Consulta:	que Ovidio escribió, en todos los medios, en CORDE
Resultado:	2 casos en 2 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias. Normal. Clasificación: Marcas:

Agrupación:

Cómo citar el CORPUS **Concordancias.**

Pantalla: 1 de 1. [Ver párrafos](#)

mí más me convence es aquella carta que Ovidio escribió a Tiberio César donde, desculpand ** un Macías, diciéndoles más dulzuras que Ovidio escribió. Pero como sintieron dél que esta **

AÑO	AUTOR
c 1550	Arce de Otá
a 1554	Anónimo

Ir arriba Pantalla: 1 de 1. [Ver párrafos](#)

5) <un ?ancto Tomás>

Consulta:	un ?ancto Tomás, en todos los medios, en CORDE
Resultado:	2 casos en 2 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias. Normal. Clasificación: Marcas:

Agrupación:

Cómo citar el CORPUS **Concordancias.**

Pantalla: 1 de 1. [Ver párrafos](#)

itosa de libros, que con un sancto Tomás y un Maestro de las Sentencias y con ** an, haciase entre ellos un Sancto Tomás y hablaba dos horas en latín -a lo me **

AÑO	AUTOR
c 1550	Arce de Otá
a 1554	Anónimo

Ir arriba Pantalla: 1 de 1. [Ver párrafos](#)

6) aquella Costanilla

Consulta:	aquella Costanilla , en todos los medios, en CORDE
Resultado:	2 casos en 2 documentos.

OBTENCIÓN DE EJEMPLOS

Recuperar Concordancias. Normal. Clasificación: Marcas:

Agrupación:

Cómo citar el CORPUS **Concordancias.**

Pantalla: 1 de 1. [Ver párrafos](#)

de Corpus Cristi aquella Costanilla y las otras calles y ventanas, que ** le donde nació, en aquella Costanilla de Valladolid, valdrían más de doc **

AÑO	AUTOR
c 1550	Arce de Otá
a 1554	Anónimo

Ir arriba Pantalla: 1 de 1. [Ver párrafos](#)

Las pesquisas que se siguieron a partir de estos datos las he venido publicando a lo largo de estos últimos 15 años y no es cuestión ahora de repetirme⁹. Diré solo que las muchas coincidencias verbales, temáticas e ideológicas entre el *Lazarillo* y el corpus de Juan Arce

9.- MADRIGAL (2008, 2014, 2019, 2020, 2022)

de Otálora (aunque en sus años de colegial se hacía llamar probablemente Juan Vélez de Otálora¹⁰) lo hacen, en mi modesta opinión, el candidato más plausible. En todo caso, el objetivo principal de este trabajo ha sido cribar un conjunto de textos a través de la informática y, tras el análisis e interpretación de datos, sacar algunas conclusiones, que son las que pongo debajo:

1. Tanto JGAAP como STYLO son instrumentos utilísimos para la clasificación de textos por autor.
2. El alto grado de acierto con textos de autoría conocida obliga a tomar en consideración las distancias que los dos programas establecen entre el *Lazarillo* y el resto de textos analizados.
3. Así, los textos jocosos atribuidos a Hurtado de Mendoza y los diálogos de Alfonso de Valdés muestran en la mayoría de los casos una distancia grande con respecto al *Lazarillo* y deben descartarse como posibles candidatos. Lo mismo pasaría con Sebastián de Horozco, aunque el texto empleado aquí sea insuficiente.
4. De todos los textos de 10.000 palabras que se han analizado en ambos programas los cuentos de Otálora presentan de manera reiterada la mayor cercanía con el *Lazarillo*, a excepción de *El asno de oro*, claro modelo del librito anónimo.
5. De todos los textos de 2.000 palabras analizados, los textos de Otálora vuelven a estar en la órbita del *Lazarillo*.
6. La separación que se observa entre el tratado V y el resto de tratados del *Lazarillo* nos hace sospechar que la composición del buldero debió ser seguramente más tardía.
7. La consistente proximidad entre el tratado V y las interpolaciones de Alcalá parece indicar un mismo agente.
8. Del mismo modo, la consistente proximidad entre los cuentos de Otálora, el tratado quinto y las interpolaciones nos lleva a pensar en una misma mano¹¹.
9. La mayoría de las pruebas ponen de manifiesto que el *Lazarillo* de 1554 y la *Segunda Parte del Lazarillo* son obras de distinto autor, pero la continuación debió escribirse dentro del círculo del autor del *Lazarillo* por la proximidad con las interpolaciones, el tratado V y los cuentos de Otálora.
10. La suma de colocaciones léxicas y paralelismos raros o exclusivos entre textos, complementada con cálculos estadísticos de n-gramas de 4 caracteres y palabras más frecuentes, parece ser una metodología eficaz en la atribución basada en datos internos.

10.– VAQUERO (40-45); MADRIGAL (2020: 213-215).

11.– *Coloquios de Palatino y Pinciano* estuvo sometido a una larga gestación con múltiples correcciones y cambios a lo largo de los años. OCASAR (2008), en un estudio magistral, señala la supresión de amplios pasajes por temor a la censura en versiones anteriores, como un pasaje extenso sobre las bulas muy en sintonía con el tratado V y, en especial, con las Interpolaciones de Alcalá.

Obras citadas

- BARAS ESCOLÁ, Alfredo (2021). «De por qué la «Novela de la tía fingida» no es de Cervantes». *Anales cervantinos* 53 (6): 293-321.
- EDER, Maciej, Jan Rybicki and Mike Kestemont (2016). «Stylometry with R: A Package for Computational Text Analysis». *The R Journal*. 8:1 (107-121).
- MADRIGAL, José Luis (2022). «El Prólogo del *Lazarillo* a la luz de los prólogos del licenciado Arce de Otálora con otras consideraciones». *Revista de Literatura Española Medieval y del Renacimiento (LEMIR)* 26: 93-124.
- (2020). «Técnicas de atribución textual en el Prólogo del *Lazarillo* y en otros textos colindantes». *Revista de Literatura Española Medieval y del Renacimiento (LEMIR)*. 24: 209-244.
- (2019). «Juan Arce de Otálora y el *Lazarillo* de 1554». *Revista de Literatura Española Medieval y del Renacimiento (LEMIR)*. 23: 323-346.
- (2014). «De nombres y lugares: el corpus del licenciado Arce de Otálora». *Revista de Literatura Española Medieval y del Renacimiento (LEMIR)* 18: 89-118
- (2008). «Notas sobre la autoría del *Lazarillo*». *Revista de Literatura Española Medieval y del Renacimiento (LEMIR)*. 12: 137-236.
- (2002). «De cómo y por qué *La tía fingida* es de Miguel de Cervantes» *Artifara*, 2. Università degli Studi di Torino.
- JUOLA, Patrick (2008). *Authorship Attribution*. Now the Essence of Knowledge: Boston.
- OCASAR ARIZA, José Luis (2008). *La lucha invisible: estudio genético-literario de los «Coloquios de Palatino y Pinciano»*, de Juan Arce de Otálora. Universidad de Valladolid
- ROSA, Javier de la, y Suárez, Juan Luis (2016). «The Life of *Lazarillo de Tormes* and of His Machine Learning Adversities: Non-traditional authorship attribution techniques in the context of the *Lazarillo*». *Revista de Literatura Española Medieval y del Renacimiento (LEMIR)*. 20: 373-438.
- SAVOY, Jacques (2020). *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*. Springer Nature.
- VAQUERO SERRANO, Carmen (2014). «Juan de Arce de Otálora: Contribución a su biografía y a la de sus familiares». *Revista de Literatura Española Medieval y del Renacimiento (LEMIR)*. 18: 9-88.

Apéndice I

1. CORPUS DE LA TÍA FINGIDA¹²

JGAAP	STYLO	palabras
ALEMAN Cuentos (Dorido y Clorinia - Dorotea)	Aleman_cuentos	10,000
CASTILLO SOLORZANO <i>Bachiller Trapaza</i>	Castillo_trapaza	10,000
CASTILLO SOLORZANO <i>La fantasma de Valencia</i>	Castillo_valencia	10,000
CASTILLO SOLORZANO <i>Las harpías de Madrid</i>	Castillo_madrid	10,000
CASTILLO SOLORZANO <i>Lisardo enamorado</i>	Castillo_enamorado	10,000
CERVANTES <i>Coloquio de los perros</i>	Cervantes_perros	10,000
CERVANTES <i>El celoso extremeño</i>	Cervantes_extremeño	10,000
CERVANTES <i>La Ilustre fregona</i>	Cervantes_fregona	10,000
CERVANTES <i>Rinconete</i>	Cervantes_rinconete	10,000
LOPE Guzmán el Bravo	Lope_bravo	10,000
LOPE <i>La desdicha por la honra</i>	Lope_honra	10,000
LOPE <i>La prudente venganza</i>	Lope_venganza	10,000
LOPE <i>Las fortunas de Diana</i>	Lope_diana	10,000
MONTALBÁN <i>La hermosa Aurora</i>	Montalban_aurora	10,000
MONTALBÁN <i>La amistad castigada</i>	Montalban_castigada	10,000
MONTALBÁN <i>La fuerza del desengaño</i>	Montalban_desengaño	10,000
MONTALBÁN <i>La villana de Pinto</i>	Montalban_pinto	10,000
PICARA JUSTINA	Picara_justina	10,000
SALAS BARBADILLO <i>La hija de la Celestina</i>	Salas_celestina	10,000
SALAS BARBADILLO <i>El sagaz Estacio</i>	Salas_estacio	10,000
SALAS BARBADILLO <i>La peregrinación sabia</i>	Salas_sabia	10,000
TIRSO <i>El bandolero</i>	Tirso_bandolero	10,000
TIRSO <i>Cigarrales</i>	Tirso_cigarrales	10,000
TIRSO <i>Los tres maridos burlados</i>	Tirso_burlados	10,000
ZAYAS <i>Aventurarse perdiendo</i>	Zayas_perdiendo	10,000
ZAYAS <i>EL castigo de la miseria</i>	Zayas_miseria	10,000
ZAYAS <i>La esclava de su amante</i>	Zayas_amante	10,000
ZAYAS <i>La inocencia castigada</i>	Zayas_inocencia	10,000

2. CORPUS DE TEXTOS SIGLO XVI (10.000)

JGAAP	STYLO	palabras
APULEYO, <i>Asno de oro</i>	Asno_oro	10,000
BUSCÓN (Quevedo, Francisco de)	Buscón	10,000

12.- Todos los textos están modernizados y la mayoría proceden de varias bibliotecas virtuales de libre acceso (por ejemplo, la Biblioteca Virtual Cervantes). El resto son transcripciones hechas por mí.

<i>Celestina</i> (Primer Acto)		10,000
CERVANTES DE SALAZAR «Carta nuncupatoria»		10,000
CERVANTES DE SALAZAR <i>Crónica de la Nueva España</i>	CdeS_CronicaNE	10,000
CERVANTES <i>Galatea</i>	Cervantes_Galatea	10,000
DELICADO Lozana andaluza	Lozana	10,000
<i>Diálogo de las transformaciones</i>		10,000
GUEVARA Epístolas familiares	Guevara_EF	10,000
GUZMÁN DE ALFARACHE (Alemán, Mateo)	Guzmán	10,000
HOROZCO Capítulo Nono Joan	Horozco_representacion	10,000
LAZARILLO	Lazarillo	10,000
LUIS DE LEON <i>De los nombres de Cristo</i>		10,000
MENDOZA <i>Carta del bachiller de Arcadia</i>	Mendoza_Arcadia	10,000
MENDOZA <i>Guerra Granada</i>	Mendoza_Granada	10,000
MENDOZA <i>Sermón de Aljubarrota</i>	Mendoza_Aljubarrota	10,000
NARVAEZ <i>Diálogo del capón</i>		10,000
OTALORA <i>Cuentos (Coloquios de Palatino y Pinciano)</i>	Otalora_cuentos	10,000
PINEDA, Juan de <i>Diálogos familiares</i>	Pineda_DF	10,000
PINEDO, Luis de, <i>Libro de Chistes</i>	Pinedo_chistes	10,000
ROJAS, Fernando de <i>Celestina</i> (sin primer acto)	Celestina	10,000
<i>Segunda Parte del Lazarillo (La)</i>	Lazarillo_cont	10,000
SILVA Feliciano <i>La segunda Celestina</i>	Silva_Celestina	10,000
<i>Tercera Celestina</i>		10,000
VALDES <i>Diálogo de Mercurio y Caronte</i>	Valdes_Mercurio	10,000
VALDES <i>Diálogo de las cosas acaecidas en Roma</i>	Valdes_roma	10,000
VALDES JUAN <i>Dialogo de la lengua</i>		10,000
VALDES JUAN <i>Doctrina cristiana</i>		10,000
<i>Viaje de Turquía</i>	Viaje_Turquia	10,000
VILLALON <i>Ingeniosa comparación</i>	Villalon_comparacion	10,000
VILLALON <i>Provechoso tratado de cambios</i>		10,000
VILLALON <i>Scholastico</i>	Villalon_scholastico	10,000

3. CORPUS DEL SIGLO XVI (2000 palabras)

JGAAP	STYLO
ASNO DE ORO	Asno_1
	Asno_2
	Asno_3
	Asno_4

BALDO	Baldo_1
	Baldo_2
	Baldo_3
	Baldo_4
CROTALON	Crot_1
	Crot_2
	Crot_3
	Crot_4
CERVANTES DE SALAZAR, <i>Crónica de la NE</i>	CdeS_CronicaNE
HOROZCO Capítulo Nono Joan	Horozco_representacion
LAZARILLO	
Primer tratado	Laz_1a
	Laz_1b
Segundo tratado	Laz_2a
	Laz_2b
Tercer tratado	Laz_3a
	Laz_3b
Quinto tratado	Laz_5
Sexto y séptimo tratado	Laz_67
Interpolaciones de Alcalá	Laz_Interp
MENDOZA <i>Carta del bachiller de Arcadia</i>	Mendoza_Bachiller1
	Mendoza_Bachiller2
MENDOZA <i>Sermón de Aljubarrota</i>	Mendoza_Aljubarrota1
	Mendoza_Aljubarrota2
	Mendoza_Aljubarrota3
	Mendoza_Aljubarrota4
OTALORA <i>Cuentos (Coloquios de Palatino y P.)</i>	Otalora_corredor
	Otalora_moriscas1
	Otalora_moriscas2
	Otalora_paraiso
	Otalora_vizcaino
PINEDA <i>Diálogos familiares</i>	Pineda_DF1
	Pineda_DF2
	Pineda_DF3
	Pineda_DF4
PINEDA (resumen de <i>Asno oro</i> en DF)	Pineda_Asno
PINEDO, Luis de, <i>Libro de Chistes</i>	Pinedo_chistes1
	Pinedo_chistes2
	Pinedo_chistes3
	Pinedo_chistes4
<i>Segunda Parte del Lazarillo</i>	Lazar_cont1

	Lazar_cont2
	Lazar_cont3
	Lazar_cont4
	Lazar_cont5
<i>Libro de las transformaciones</i>	Trans_1
	Trans_2
VALDES <i>Diálogo de Mercurio y Caronte</i>	Valdes_Mercurio1
	Valdes_Mercurio2
	Valdes_Mercurio3
	Valdés_Mercurio4
VALDES <i>Diálogo de las cosas acaecidas en Roma</i>	Valdes_roma1
	Valdes_roma2
	Valdes_roma3
	Valdes_roma4
<i>Viaje de Turquía</i>	Viaje_Turquia1
	Viaje_Turquia2

Apéndice II

KNN con distancia métrica Hellinger (K-Nearest Neighbor Driver with metric Hellinger Distance)

LAZARILLO 10.000

C4	W
1. MENDOZA <i>Sermón de Aljubarrota</i> 1.03125	1. VALDES (<i>Mercurio</i>) 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. <i>Segunda Parte Lazarillo</i> 1.125
4. OTALORA <i>Cuentos</i> 1.25	4. OTALORA <i>Cuentos</i> 1.25
5. <i>Asno de oro</i> 1.5	5. <i>Asno de oro</i> 1.5

LAZARILLO 2.000

Laz_1a

C4	W
1. MENDOZA <i>Sermón de Aljubarrota</i> 1.03125	1. <i>Lozana andaluza</i> 1.03125
2. <i>Segunda Parte Lazarillo</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. CERVANTES <i>Galatea</i> 1.125	3. <i>Segunda Parte Lazarillo</i> 1.125
4. OTALORA <i>Cuentos</i> 1.25	4. OTALORA <i>Cuentos</i> 1.25
5. <i>Asno de oro</i> 1.5	5. <i>Asno de oro</i> 1.5

Laz_1b

C4	W
1. <i>Diálogo de las transformaciones</i> 1.03125	1. VALDES (<i>Mercurio</i>) 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. <i>Asno de oro</i> 1.125
4. OTALORA Cuentos 1.25	4. <i>Segunda Parte Lazarillo</i> 1.25
5. <i>Asno de oro</i> 1.5	5. OTALORA Cuentos 1.5

Laz_2a

C4	W
1. <i>Diálogo de las transformaciones</i> 1.03125	1. <i>Viaje Turquía</i> 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. OTALORA Cuentos 1.125
4. <i>Asno de oro</i> 1.25	4. <i>Segunda Parte Lazarillo</i> 1.25
5. OTALORA Cuentos 1.5	5. <i>Asno de oro</i> 1.5

Laz_2b

C4	W
1. <i>Diálogo de las transformaciones</i> 1.03125	1. VALDES (<i>Mercurio</i>) 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. CERVANTES <i>Galatea</i> 1.0625
3. OTALORA Cuentos 1.125	3. <i>Asno de oro</i> 1.125
4. <i>Asno de oro</i> 1.25	4. OTALORA Cuentos 1.25
5. <i>Segunda Parte Lazarillo</i> 1.5	5. <i>Segunda Parte Lazarillo</i> 1.5

Laz_3a

C4	W
1. CERVANTES <i>Galatea</i> 1.03125	1. <i>Viaje Turquía</i> 1.03125
2. <i>Viaje Turquía</i> 1.0625	2. <i>Lozana andaluza</i> 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. <i>Segunda Parte Lazarillo</i> 1.125
4. OTALORA Cuentos 1.25	4. OTALORA Cuentos 1.25
5. <i>Asno de oro</i> 1.5	5. <i>Asno de oro</i> 1.5

Laz_3b

4 n-gramas de caracteres C4	2 n-gramas de palabras W
1. <i>Viaje Turquía</i> 1.03125	1. <i>Viaje Turquía</i> 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. <i>Lozana andaluza</i> 1.0625
3. OTALORA Cuentos 1.125	3. <i>Asno de oro</i> 1.125
4. <i>Segunda Parte Lazarillo</i> 1.25	4. <i>Segunda Parte Lazarillo</i> 1.25
5. <i>Asno de oro</i> 1.5	5. OTALORA Cuentos 1.5

Laz_5

C4	W
1. MENDOZA <i>Sermón de Aljubarrota</i> 1.03125	1. MENDOZA <i>Sermón de Aljubarrota</i> 1.03125
2. CERVANTES <i>Galatea</i> 1.0625	2. VALDES (<i>Mercurio</i>) 1.0625
3. <i>Asno de oro</i> 1.125	3. <i>Asno de oro</i> 1.125
4. <i>Segunda Parte Lazarillo</i> 1.25	4. <i>Segunda Parte Lazarillo</i> 1.25
5. OTALORA Cuentos 1.5	5. OTALORA Cuentos 1.5

Laz_interp

4 n-gramas de caracteres C4	2 n-gramas de palabras W
1. CERVANTES <i>Galatea</i> 1.03125	1. C DE SALAZAR CNE 1.03125
2. C DE SALAZAR CNE 1.0625	2. VALDES (<i>Mercurio</i>) 1.0625
3. <i>Segunda Parte Lazarillo</i> 1.125	3. <i>Asno de oro</i> 1.125
4. <i>Asno de oro</i> 1.25	4. <i>Segunda Parte Lazarillo</i> 1.25
5. OTALORA Cuentos 1.5	5. OTALORA Cuentos 1.5

Apéndice III

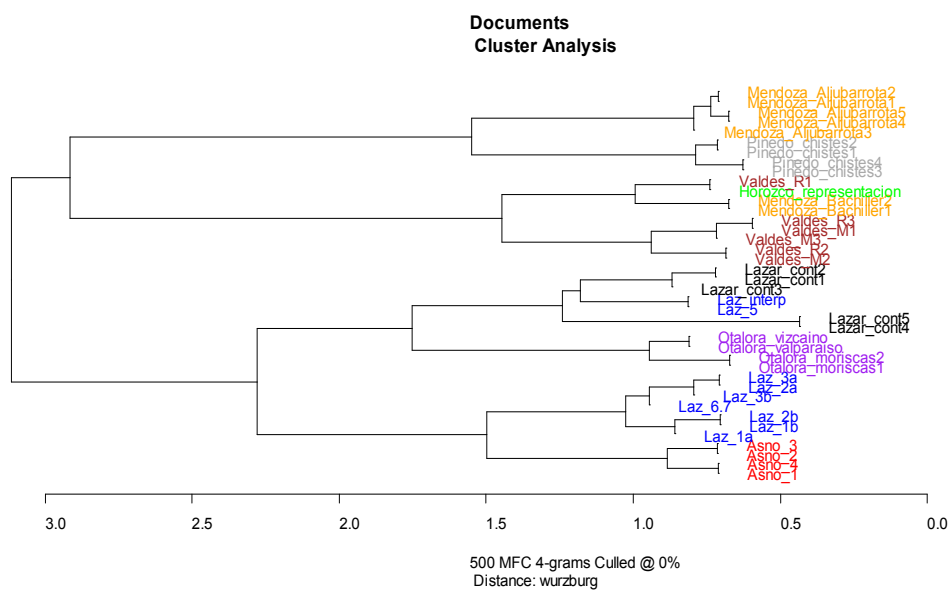
CORPUS (2000 palabras)

500 MFC 4-grams

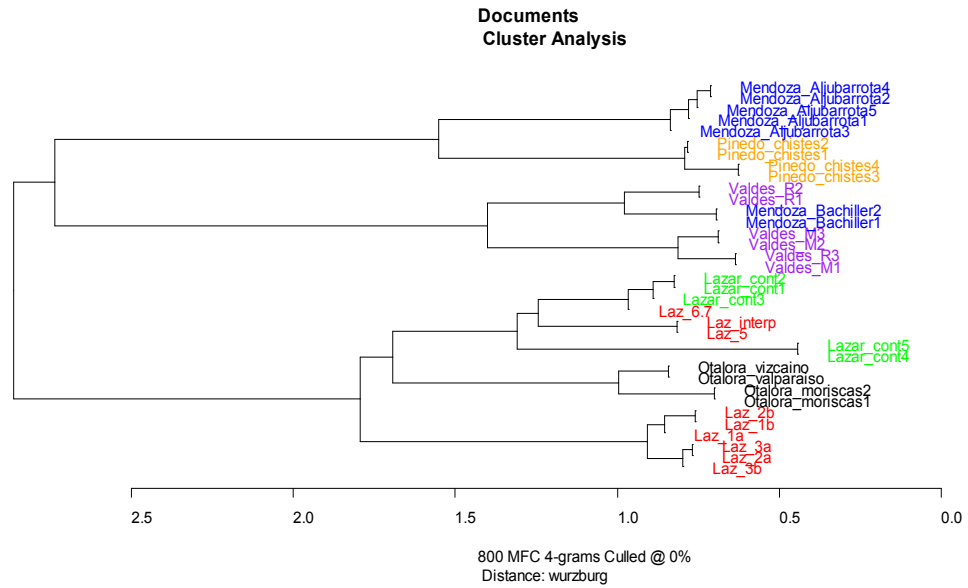
Distance wurzburg

1. ANÁLISIS DE CONGLOMERADOS:

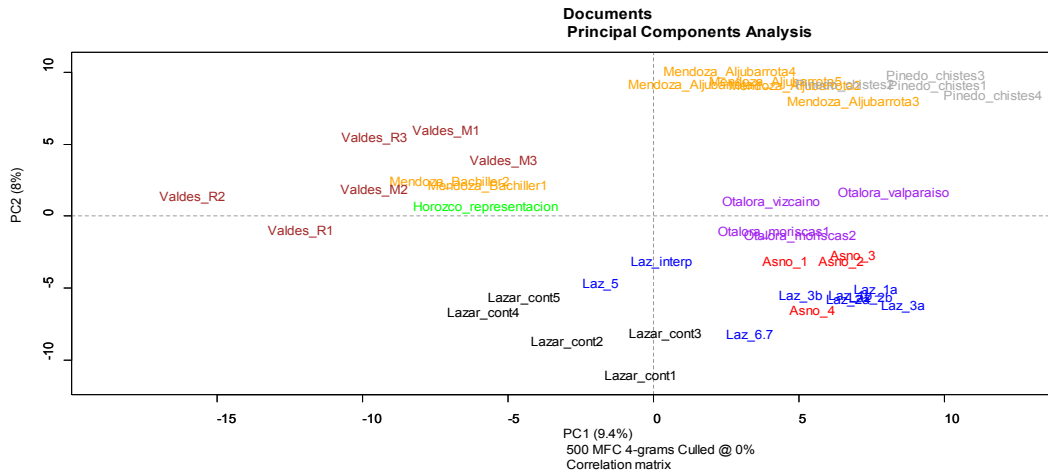
500 palabras

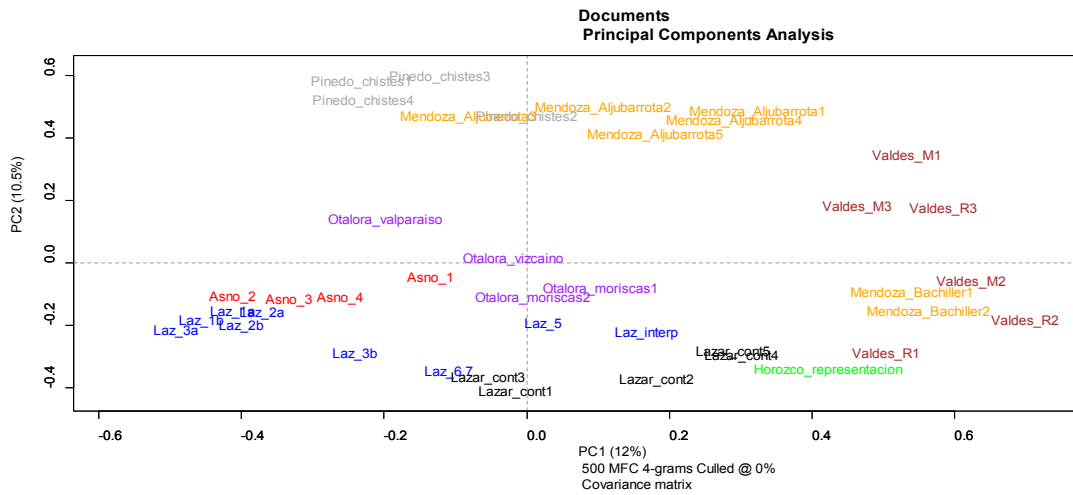
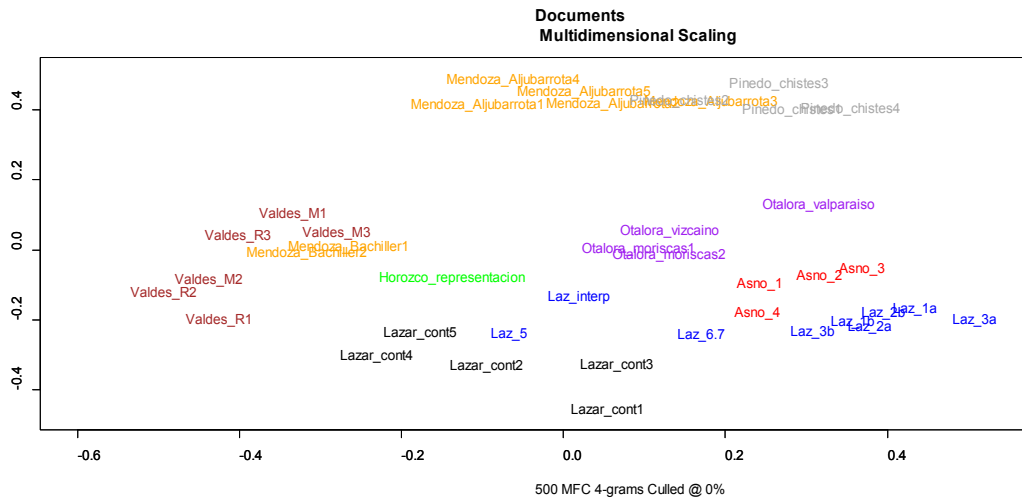


800 palabras (sin *Asno de oro*)

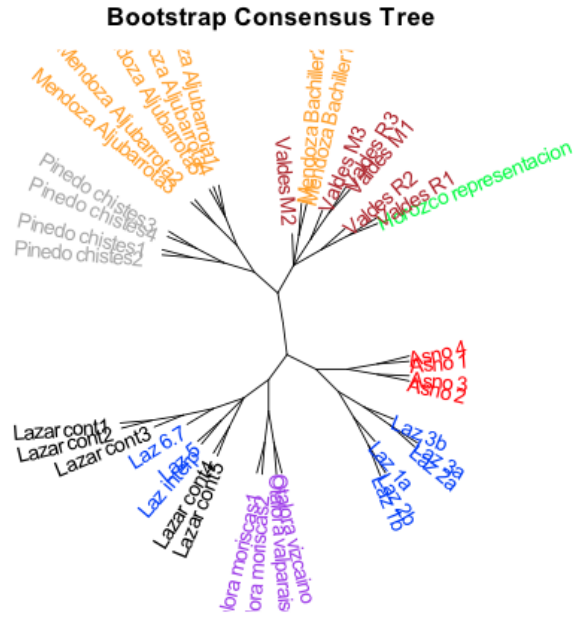


2. ANÁLISIS DE COMPONENTES PRINCIPALES (PCA):

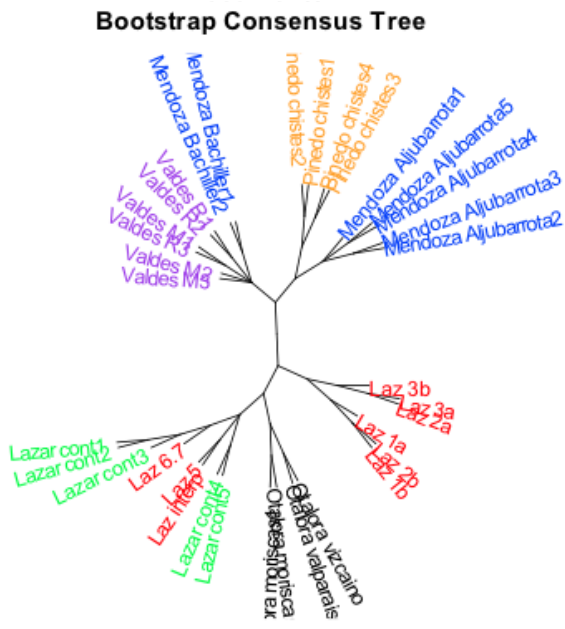




3) BOOTSTRAP CONSENSUS TREE (100-1000 MFC 4-grams).



100-1000 MFC 4-grams Culled @ 0%
Distance: wurzburg Consensus 0.5



100-1000 MFC 4-grams Culled @ 0%
Distance: wurzburg Consensus 0.5

