

## La novela ejemplar *La tía fingida*, atribuida a Cervantes


Alfredo Rodríguez López-Vázquez  
Universidad de La Coruña

### Resumen

En este estudio se analiza el estado de la cuestión sobre la autoría de la novela breve *La tía fingida*, siguiendo las dos versiones textuales de la obra, la del manuscrito Porrás de la Cámara y la del Códice de la Biblioteca Colombina de Sevilla. Aplicando un aparato matemático basado en matrices lineales, logaritmos y ecuaciones basadas en el número  $e$  (2,718) estudiamos 20 índices textuales de esta novela en la obra de Cervantes, Quevedo, Vicente Espinel, López de Úbeda y Salas Barbadillo. Concluimos que la obra es de Miguel de Cervantes con una probabilidad superior al 95 por ciento.

### Abstract

In this study we analyze the state of the art about the authorship of the novella *La tía fingida*, following its two textual versions, that of the Porrás de la Cámara manuscript and that of the Codex in the Biblioteca Colombina of Seville. Applying a mathematical apparatus based on linear matrixes, logarithms, and  $e(2.718)$ -based equations we study 20 textual indexes of this novella in the works by Cervantes, Quevedo, Vicente Espinel, López de Úbeda and Salas Barbadillo. We conclude that the piéce belongs to Miguel de Cervantes with a probability superior to 95 percent.

 Los avatares editoriales de *La tía fingida* y su atribución a Cervantes han ocupado, con dispar fortuna, los dos últimos siglos, desde la edición de 1818, preparada por Martín Fernández de Navarrete siguiendo el manuscrito del racionero sevillano Francisco Porrás de la Cámara, que copia esta obra para solaz y lectura del arzobispo Fernando Niño de Guevara, acompañada de otras dos novelas: *Rinconete y Cortadillo* y *El celoso extremeño*. La edición de Navarrete, conocida como edición Francesón-Wolf, corrige gran cantidad de errores y deturpaciones<sup>1</sup> de la primera, hecha por el erudito Isidoro Bosarte, a partir del manuscrito Porrás. Lamentablemente este manuscrito desapareció en 1823, con lo que hoy en día sólo disponemos de la edición hecha por Navarrete, la edición perpe-

<sup>1</sup> «El señor Arrieta sacó una copia de aquella novela, que, con algunas mutilaciones publicó en nuestros días». *Obras de Miguel de Cervantes Saavedra*, BAE, Madrid: Rivadeneyra, 1864, p. xxviii.



trada por Bosarte y la posterior de Arrieta<sup>2</sup>. A cambio, el descubrimiento de otra versión textual de la novela en un códice de la Biblioteca Colombina, a cargo de Aureliano Fernández-Guerra en 1845, nos proporciona un nuevo texto, idéntico en los contenidos narrativos, pero con interesantes modificaciones formales, semejantes a las que podemos entrar en los casos de *Rinconete y Cortadillo* y en *El celoso extremeño*, si comparamos la copia de estas dos novelas ejemplares hechas por Porrás de la Cámara, con la versión final preparada por Cervantes en su edición de 1613. Una primera hipótesis crítica apunta a que el texto de la Colombina es posterior al texto de la copia Porrás, probablemente trasladado hacia 1605, recién publicada la primera parte del *Quijote*. Conviene no descartar, además, que en los casos en que el texto de la Colombina presenta omisiones respecto al texto Porrás, podrían deberse tanto a supresión consciente de Cervantes como a mutilaciones en el proceso de copia.

Desde el siglo XIX hasta muy recientemente se han expuesto distintas argumentaciones para apoyar ambas posturas. Las podemos resumir en los nombres de sus principales valedores, Julián Apráiz, favorable a la atribución cervantina, y Francisco de Icaza, contrario a ella. A mediados del siglo XX, Valbuena Prat, que incluye *La tía fingida* en su edición de obras de Cervantes, apunta que «aún hoy conserva su fuerza la argumentación y detalladísimo estudio de un notable cervantista, Julián Apráiz»<sup>3</sup>. A comienzos del siglo XXI la polémica subsiste y aunque las últimas aportaciones críticas de J.L. Madrigal (2002) y de F. López (2011) refuerzan la atribución cervantina, sigue sin proponerse un estudio capaz de aportar elementos concluyentes para esta atribución, de modo que un reciente editor de las *Novelas ejemplares* se limita a apuntar, hablando del manuscrito Porrás de la Cámara, que «traía también una novela de autor ignoto, *La tía fingida*, de ambientación celestinesca y prostibular, cuya atribución a Cervantes constituye un aspecto polémico desde su descubrimiento»<sup>4</sup>. Sin duda el aspecto polémico alude al fervor con que han sido expuestas las dos hipótesis o conjeturas sobre la atribución: que la obra sea de Cervantes (Apráiz) o que no lo sea (Icaza). La cosa se complica al existir dos fases distintas de elaboración textual en dos manuscritos transmitidos con desigual fortuna. El manuscrito Porrás ha sido transmitido con censuras y deturpaciones por Bosarte y por Arrieta y posteriormente, tal vez de forma fidedigna, por Fernández Navarrete antes de su pérdida, con lo que sólo podemos acudir al texto Francesón-Wolf de 1818, que Navarrete edita modernizando la ortografía conforme a los usos del siglo XIX<sup>5</sup>. A cambio, sí disponemos del

<sup>2</sup> «Posteriormente ha publicado D. Agustín de Arrieta, al fin del *Espíritu de Miguel de Cervantes* (un tomo en 8º) esta novela, pero no sacada del original, sino de otra copia poco exacta, y suprimiendo por la decencia pública varios diálogos y pasajes graciosos y oportunos», puntualiza Navarrete en el apéndice a la edición Francesón-Wolf (p. 35).

<sup>3</sup> M. Cervantes, *Obras Completas*, Tomo II, Madrid, Aguilar, 1940, p. 1107.

<sup>4</sup> M. Cervantes, *Novelas ejemplares*, Barcelona, Ed. Crítica, 2010 (2ª ed.), p. 70. Ed. Jorge García López.

<sup>5</sup> «Esta copia se hizo por una del señor Estala, y en la confrontación con el original han resultado las enmiendas y correcciones que van anotadas: siendo de advertir además para mayor exactitud, que el original dice siempre *muncha*, *parecer*, *codicia*, *santenuffio*, *bajaranos*, *dubda*, *puncto*, etc. Por *mucha*, *parecer*, *codicia*, *duda*, etc.», Navarrete, p. 35.



manuscrito de la Colombina (sign. 56-4-34, pp. 77-88), que representa otro estado del texto que no puede proceder de una copia, ni directa, ni diferida, del manuscrito Porras. Hay, pues, una doble vía de transmisión a partir de un texto original sobre cuya autoría hay una polémica que tiene a Cervantes como centro.

La propuesta metodológica que elaboramos aquí trata de dilucidar por medios objetivos, verificables y, si es el caso, refutables las bases de esta atribución cervantina que hasta ahora ha transitado en bastantes ocasiones por la vía de la inuición ocasional no siempre fundada, de las opiniones críticas, las convicciones personales y las creencias subjetivas, lo que explica la carga polémica de las distintas posturas en el debate. A partir de argumentaciones más o menos sólidas el fondo del debate ha sido esencialmente ideológico. Esta propuesta se basa en el escrutinio de secuencias textuales de *La tía fingida* y su verificación de uso en Cervantes y otros autores coetáneos. Se trata de comprobar en qué autor o autores aparecen esas secuencias, con qué frecuencia lo hacen y qué valor discriminante tienen los resultados obtenidos. Está claro que una secuencia que aparezca en Cervantes, pero también cinco o seis autores más de la época tiene muy poca relevancia crítica. Salvo que en Cervantes aparezca de forma constante y en los demás autores ocasionalmente, en cuyo caso habría que encontrar una fórmula para determinar su valor como índice autoral que incluyera la frecuencia de uso en cada autor, el número total de usos en un período dado y un índice que relacionara el número de elementos de la secuencia con su probabilidad de aparición.

El primer filtro que llevamos a cabo tiene que ver con el principio de uso restringido: sólo se analizan secuencias textuales usadas por un máximo de cinco autores en el período de 1600 a 1620, de acuerdo con la documentación que registra el CORDE. Esto garantiza que en el corpus general de autores de ese período la secuencia analizada es poco frecuente. En ocasiones, ese máximo de cinco autores se reduce a un solo autor, con lo que el valor discriminatorio es máximo.

Pondré un ejemplo del método acudiendo a la primera unidad narrativa de *La tía fingida* (en adelante TF) según el texto Porras:

{Pasando por cierta calle de Salamanca, dos estudiantes manchegos y mancebos, más amigos del *baldeo* y *rodancho* que de Bártulo y Baldo, vieron en una ventana de una casa y tienda de carne una celosía y pareciéndoles novedad, porque la gente *de la tal casa*, si no se descubría y apregonaba, no se vendía, queriéndose informar del caso deparoles su diligencia un oficial vecino, pared en medio, el cual les dijo:}

Esta unidad narrativa consta de setenta palabras. Una secuencia como 'pared en medio', que parece prometedora como índice, resulta no tener valor discriminante, ya que aparece 25 veces en veinte documentos de ese período. La usa Cervantes, pero también la usan Lope, Tirso y una docena más de autores. En cambio los términos de germanía 'baldeo' y 'rodancho', aparecen en ese período usados catorce veces en cinco documentos y nueve veces en seis documentos, con lo que su valor discriminante es claro. Sobre todo porque casi todos esos usos aparecen en el *Romancero de germanía* de Juan Hidalgo (1609), de modo que la mayor parte



de los usos registrados son anónimos. Entre los pocos autores que usan esos vocablos de germanía tenemos a Cervantes, Quevedo y al autor de *La pícaro Justina*, lo que acota el elenco de autores con mucha precisión. Dado que el romancero de Hidalgo es de 1609 y el manuscrito *Porras* de la Cámara probablemente es de 1605, está claro que el uso de esos términos antes de la publicación del romancero de Hidalgo es muy significativo.

Más interesante es la secuencia textual 'de la tal casa', porque establece los límites de elementos en que se aprecia valor discriminante. No tiene sentido buscarlo en el vocablo 'casa', del que el CORDE registra 17.981 usos en ese período; pero sí tiene más interés el modismo 'tal casa', que aparece sólo veintidós veces, de las cuales ocho están en San Juan Bautista de la Concepción, tres en Alonso de Ledesma, dos en Luque Fajardo y una en Cervantes. Si filtramos el elemento anterior del sintagma, la secuencia 'la tal casa' nos deja sólo un elenco de cuatro autores en un total de siete registros. El siguiente filtro, de la secuencia entera 'de la tal casa' sólo aparece en un autor: Miguel de Cervantes. Así la secuencia de tres vocablos tiene ya valor discriminante; la de los cuatro es excluyente, ya que selecciona sólo a Cervantes. Vale la pena cotejar ese comienzo de TF según el manuscrito *Porras* con la variante textual que presenta el Códice Colombino:

#### ibéricas y latinoamericanas

{Pasando por una calle de Salamanca dos estudiantes mancebos, más amigos del *baldeo* o *rodancho* que de Bártulo o Baldo, alzaron acaso los ojos a una ventana y vieron en ella una celosía puesta, que otras veces no habían visto, y pareciéndoles cosa nueva, considerando qué novedad era aquella, porque ellos sabían que en aquella casa no vivía gente que requiriese poner celosías en las ventanas, queriéndose informar de un vecino oficial, que pared en medio estaba, el cual les dijo:}

Probablemente en la copia manuscrita de la Colombina falta un sintagma que el amanuense ha omitido: 'queriéndose informar {del caso} de un vecino oficial'. En todo caso hay algo evidente: en este segundo texto hay 81 palabras en vez de setenta y, si se admite la omisión de 'del caso', 83. En este caso ya no aparece el sintagma excluyente 'de la tal casa', lo que podría hacer sospechar que las modificaciones del texto no se deben a la pluma original. Sin embargo el análisis de variantes nos vuelve a llevar a Cervantes, en la secuencia 'alzaron acaso los ojos'. La secuencia 'acaso los ojos' sólo registra tres usos en ese período. Dos de los usos están en Cervantes y uno en Salas Barbadillo. Pero el contexto selecciona de nuevo a Cervantes que, en ambos casos, presenta la secuencia precedida del verbo 'alzar' ('alcé acaso los ojos/ alzó acaso los ojos'), frente a Salas, que lo usa precedido del verbo 'volver'. La secuencia 'ALZ-' + 'acaso los ojos' es privativa de Cervantes y además está repetida en su obra.

¿Hay algún otro estilema que tenga valor discriminante dentro de la variante textual de la Colombina? Hay, en efecto, uno y esta vez no apunta a Cervantes, sino a otro autor muy a considerar: Vicente Espinel. Se trata de la construcción 'considerando qué novedad era aquella', que podemos subdividir en dos





secuencias: a) 'considerando qué' y b) 'qué novedad era aquella'. La primera secuencia, usando el gerundio del verbo 'considerar', aparece 8 veces en el CORDE en ese período, pero tan sólo la usan cuatro autores, ya que Espinel la usa 3 veces y San Juan Bautista de la Concepción, dos veces. En realidad, si en vez del gerundio estricto ampliamos al esquema 'CONSIDER- + qué' (como hemos hecho con 'ALZ- +acaso los ojos'), esta construcción deja de tener valor significativa, ya que aparece en varios autores más. Sin embargo la segunda secuencia, 'qué novedad era aquella' resulta muy interesante porque volvemos a ver su uso en Vicente Espinel, en un elenco ya muy reducido. La fórmula aparece en Mármol Carvajal, Vicente Espinel y en un anónimo, por lo tanto, en menos de cinco autores. Hay que hacer notar que estamos hablando de la variante textual de la *Colombina*, que es una copia posterior a 1631, como ha detectado Luis Astrana Marín; esto quiere decir que se podría proponer a Espinel como posible modificador de un texto inicial ajeno. Pero conviene introducir a Espinel dentro del elenco de autores cotejables con Cervantes, que en principio sigue presentando más índices discriminantes de secuencias de cuatro vocablos ('de la tal casa' y 'ALZ- acaso los ojos') y con un uso más excluyente.

En resumen, disponemos de cinco índices asumiendo las dos variantes textuales de *Porras* y *Colombina*, y disponemos también cinco autores: Cervantes, Quevedo, Espinel, el autor de *La pícaro Justina*<sup>6</sup> y Salas Barbadillo, por este orden en la disposición de las filas. Esto permite trabajar con una matriz cuadrada de rasgos distintivos, en donde las filas corresponden a esos cinco autores y las columnas a los cinco rasgos que hemos detectado como discriminantes. Los rasgos son: a<sub>1</sub> {baldeo}, a<sub>2</sub> {rodancho}, a<sub>3</sub> {de la tal casa}, a<sub>4</sub> {ALZ-+acaso los ojos}, a<sub>5</sub> {qué novedad era aquella}. La matriz aumentada con una columna suplementaria indica el número total (T) de elementos de cada autor. Asignando los valores [0,1] a la presencia o ausencia de rasgos, la matriz resultante es la siguiente:

$$\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & T \\ 1 & 0 & 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Asumiendo la presencia o ausencia de rasgos, sin ponderar, Cervantes es el autor más coincidente para el conjunto {a<sub>1</sub>...a<sub>5</sub>}. La ponderación de rasgos incluyendo una fórmula que desarrolle el número de palabras, el número de autores que usan el elemento y el número de veces que la usa un autor amplía considerablemente la relevancia de Cervantes, porque los rasgos {a<sub>3</sub>, a<sub>4</sub>, a<sub>5</sub>} son secuencias de cuatro términos, mientras que los dos primeros rasgos son de un vocablo y los usan varios autores. No es el momento ahora de desarrollar esa fór-

<sup>6</sup> La polémica sobre la atribución de esta obra está sin resolver. Conforme a la reciente edición de David Mañero mantenemos el nombre clásico de Francisco López de Úbeda.



mula, pero sí de hacer ver que un tratamiento cuidadoso, que combine matrices y fórmulas algebraicas permite ahondar y afinar los resultados iniciales.

En función de ello veamos ahora los restantes índices que podemos encontrar en el macrotexto formado por las dos variantes textuales, los manuscritos P y C. Mantenemos el elenco de autores inicial para el resto de los elementos analizados.

La segunda unidad narrativa presenta también variantes textuales de interés. El manuscrito Porras da el siguiente texto:

{«Señores, *habrá ocho días que vive en esta casa una señora forastera, medio beata y de mucha autoridad. Tiene consigo una doncella de extremado parecer y brío, que dicen ser su sobrina. Sale con un escudero y dos dueñas, y según he juzgado es gente honrada y de gran recogimiento: hasta ahora no he visto entrar persona alguna de esta ciudad ni de otra a visitallas, ni sabré decir de cuál vinieron a Salamanca. Mas lo que sé es que la moza es hermosa y honesta, y que el fausto y autoridad de la tía no es de gente pobre.*»

La relación que dio el vecino oficial a los estudiantes les puso codicia de *dar cima* a aquella aventura, porque siendo pláticos en la ciudad y deshollinadores de cuantas ventanas tenían albahacas con tocas, en toda ella no sabían que tal tía y sobrina hubiese cursantes en su Universidad, principalmente que viniesen a vivir a semejante casa, en la cual, por ser buen peaje, siempre se había vendido tinta, aunque no de la fina; que hay casas, así en Salamanca como en otras ciudades, que llevan de suelo vivir siempre en ellas mujeres cortesanias, y por otro nombre trabajadoras o enamoradas.}

El primer índice importante ( $b_1$ ) es la secuencia 'habrá ocho días que': el CORDE tan sólo registra un ejemplo de este uso y está en *Rinconete y Cortadillo*; se trata, pues, de otro índice excluyente, relacionado además con una de las novelas ejemplares copiadas por Porras de la Cámara. También la expresión 'sabré decir de' ( $b_2$ ), que sólo aparece una vez en el CORDE; está en el *Persiles* de Cervantes, en 1616. El tercer índice ( $b_3$ ) es 'lo que sé es que', coloquialismo usado por el oficial y que sólo se registra en dos autores: Cervantes y Fernández de Avellaneda. Cervantes la usa también en *Rinconete y Cortadillo*, lo que es, de nuevo, llamativo; aparece también en el *Persiles*. El caso de Avellaneda, habilísimo suplantador del estilo cervantino evidencia que este coloquialismo es típico de Cervantes. El cuarto ejemplo ( $b_4$ ) es 'dar cima', que aparece registrado una sola vez entre 1600 y 1620. De nuevo es Cervantes el único autor que la usa. De hecho, la usa en el mismo entorno que incluye 'aventura'. El último de esta segunda unidad narrativa, ( $b_5$ ) resulta un poco sorprendente, porque el sintagma 'dos dueñas' podría pensarse como uno de uso frecuente o habitual entre los narradores de la época. Sin embargo no es así: aparece tan sólo ocho veces, de las cuales tres están en Cervantes (dos veces en *La ilustre fregona* y una en la segunda parte del *Quijote*) y dos veces en Quevedo. Lo usan también, una sola vez, Cabrera de Córdoba, Prudencio de



Sandoval y Avellaneda, que no están en el elenco estudiado. Sin necesidad de volver a visualizar esto en una matriz de rasgos, podemos sintetizarlo con una adscripción de elementos por cada autor:

Cervantes {b<sub>1</sub>, b<sub>2</sub>, b<sub>3</sub>, b<sub>4</sub>, b<sub>5</sub>}  
 Quevedo {b<sub>5</sub>}

Espinel, López de Úbeda y Salas Barbadillo no usan ninguno de estos cinco índices. Ya hemos visto que Alonso Fernández de Avellaneda, en este sentido, es mejor candidato que Quevedo, ya que usa dos de estos cinco índices. Parece evidente que en su caso se trata de 'huellas de lectura', lo que avala todavía más que el repertorio que seleccionamos es muy significativo. En adelante me limitaré a detallar el resto de los índices asignando la referencia, los autores que los usan y en el caso de Cervantes los textos en donde aparecen. Señalaré también los casos en los que índice se encuentra sólo en el Códice de la Colombina.

(c<sub>1</sub>) «o que no comían en ella *sus moradoras*». Este sintagma sólo se registra una vez en el CORDE: en *El licenciado Vidriera*: «vio que estaban a la puerta della muchas de sus moradoras».

(c<sub>2</sub>) «con *unas tocas blancas* como la nieve». Esta secuencia de cuatro vocablos solo se registra en Cervantes. Está repetida en la segunda parte del *Quijote*: «con unas tocas blancas de delgado canequí» y «con unas tocas blancas, repulgadas y luegas».

(c<sub>3</sub>) «de lo cual *coligieron*, o que no comían en ella». Tan sólo hay un ejemplo de uso en el CORDE de esta secuencia. Está en Cervantes, en la segunda parte del *Quijote*: «de lo cual coligieron todos que de consentimiento».

(c<sub>4</sub>) «no le *salió vano su pensamiento*». Esta secuencia de seis vocablos sólo aparece en el texto de la Colombina, ya que la variante del manuscrito Porras tiene «y no le salió vana su presunción». El hecho de que el único ejemplo de esta secuencia aparezca en una obra de Cervantes, *El celoso extremeño* que además es una de las novelas copiadas en el manuscrito Porras apunta a que el propio amanuense modificó el giro cervantino o bien Cervantes corrigió 'vana su presunción' en 'vano su pensamiento'. El sintagma 'su presunción' aparece ocho veces en el CORDE y el único autor que lo repite es Cervantes en la primera parte del *Quijote* y en el *Viaje del Parnaso*. Lo usan también Vicente Espinel y Salas Barbadillo. El índice cervantino significativo es 'vano su pensamiento'.

(c<sub>5</sub>) «más largas que *una sobrepelliz* de un canónigo portugués». El sintagma 'una sobrepelliz' aparece registrado 31 veces, pero 29 son en documentos anónimos. El único autor que lo usa es Vicente Espinel, dos veces. El vocablo sin el determinante lo usan varios autores más, entre ellos Quevedo.

De este modo, de los cinco índices, cuatro están en Cervantes y uno en Espinel. En los casos de Cervantes, con secuencias de dos, cuatro y seis vocablos, la aplicación de una fórmula que pondere los resultados refuerza notablemente la atribución cervantina.



(d<sub>1</sub>) «con *un gran rosario* al cuello». Tan sólo tres usos en el CORDE: Cervantes, Salas Barbadillo y Avellaneda. Cervantes en la primera parte del *Quijote*.

(d<sub>2</sub>) «*rosario* al cuello *de cuentas* sonadoras». La construcción ‘rosario de cuentas’, que aquí tiene intercalado un sintagma, aparece también una sola vez en el CORDE y de nuevo está en Cervantes, en la segunda parte del *Quijote*.

(d<sub>3</sub>) «de los del *tiempo del Conde Fernán González*». La secuencia completa de cinco vocablos sólo la usa Salas Barbadillo.

(d<sub>4</sub>) «con su *sayo de velludo*, ya sin vello». Aparece siete veces, pero sólo la usan tres autores. En Cervantes está repetido y en Quevedo aparece una sola vez. Los otros cuatro casos son de Silva y Figueroa, que no está en el elenco. El índice tiene interés por el juego de ingenio ‘velludo ya sin vello’.

(d<sub>5</sub>) «*capa de fajas*». Este sintagma sólo aparece cinco veces, pero tres de ellas de autor anónimo. Los dos autores que lo usan son Cervantes, en la primera parte del *Quijote* y el autor del *Guzmán apócrifo*.

De nuevo vemos en este repertorio de cinco índices que Cervantes presenta cuatro, frente a Salas Barbadillo, que tiene dos. Ni Quevedo, ni Vicente Espinel, ni López de Úbeda usan estos índices.

El resultado global de estos veinte primeros índices parece suficientemente significativo. Sin hacer todavía la ponderación según variables de autores, usos particulares y número de vocablos en el índice, Cervantes presenta dieciséis, es decir, un ochenta por ciento del total estudiado; de los otros cuatro autores, Salas Barbadillo aparece con tres y Quevedo, Espinel y López de Úbeda con dos cada uno. El índice procede de una cala exhaustiva y parece suficiente como para atribuir la obra a Cervantes tanto en su primera redacción (manuscrito Porras) como en la fase tardía del códice de la Colombina. Veamos ahora la cuestión de la fórmula ponderada para aquilatar las cifras que hemos obtenido inicialmente. Esta fórmula la hemos propuesto ya para analizar la pertinencia de la atribución cervantina de *La conquista de Jerusalén por Godofre de Bullón* y también para descartar la atribución del *Auto de la Virgen de Guadalupe*, que parece más bien obra de la primera época de Lope de Vega, en torno a 1590.

Se trata de integrar en una única fórmula las variables ‘número de palabras’ (s, por ‘secuencia’), ‘número de veces que aparece la expresión en un autor’ (subíndice n) y ‘número de autores que usan la expresión (subíndice a, por ‘autor’). El coeficiente de variación se expresa como C y la fórmula que integra todo esto es la siguiente:

$$C = e^{s-1} \cdot \frac{1 + \text{Log}_n}{1 + \text{Log}_a}$$

Lo primero que hay que explicar, ya que no es frecuente en la literatura crítica filológica, es la función del número *e* y el porqué de utilizar logaritmos neperianos y no logaritmos decimales. El número *e*, cuyo valor aproximado es 2,718, es precisamente la base de los logaritmos neperianos, que explican una gran cantidad





de fenómenos naturales y sociales, del mismo modo que 10 es la base de los logaritmos decimales y de la forma habitual de computar magnitudes a gran escala. De esta forma una cifra como 1.492 se expresa como la suma de  $10^3 + 4 \cdot 10^2 + 9 \cdot 10^1 + 2 \cdot 10^0$ . En este caso estamos usando la potenciación y la base 10. El uso de logaritmos corresponde a la operación matemática inversa a la potenciación, de modo que si  $10^3$  corresponde a la magnitud 1.000, en consecuencia el logaritmo del número 1.000, mientras usemos la base 10 está dado por el superíndice 3.

Sucede que utilizar la base 10 para tratar la información de magnitudes que habitualmente varían entre 1 y 25 (los usos de vocablos en autores y períodos dados) resulta inútil, innecesario o insuficiente, ya que sólo podríamos clasificar dos categorías, las magnitudes entre 1 y 9, que tienen 0 en la característica (la parte entera, antes de la parte decimal) del logaritmo, y las magnitudes entre 10 y 15, que tienen característica 1 en su logaritmo. Para este viaje podíamos trabajar también con el modelo binario de presencia/ausencia, con el agravante que anularía la notable diferencia de que un término aparezca una vez o que aparezca nueve veces, o que aparezca en un solo autor o en ocho. El recurso a los logaritmos neperianos, de base  $e$  (2,718) facilita el trabajo analítico enormemente, al diferenciar cuatro categorías según la magnitud sea inferior a superior a  $e$ ,  $e^2$ ,  $e^3$  y  $e^4$ . Esas categorías corresponden a usos **ocasionales** (una o dos veces), **repetidos** (entre tres y siete veces), **frecuentes** (entre ocho y veinte veces) y **característicos** (entre 21 y 53 veces). Un ejemplo espectacular de la necesidad de hacer entrar en la tipología los casos superiores a veinte usos en un autor lo tenemos con la palabra 'cañafístola', que conocen de sobra todos los lectores de Gonzalo Fernández de Oviedo. El CORDE (11/6/2013) registra el uso de esta palabra 54 veces entre 1525 y 1557. Entre los cronistas de Indias la diferencia de su uso es muy notable: Oviedo tiene 35 registros (**característico**); López de Gómara y Bartolomé de las Casas, tres y cuatro (**repetido**) y Motolinía, dos (**ocasional**).

El hecho de usar el superíndice (s-1) como exponente de la base  $e$  resulta esencial para garantizar un módulo común a las variaciones de uso de autor y de número de autores. Si estamos analizando un solo vocablo (en TF: 'baldeo' o 'rodancho', por ejemplo), el exponente (s-1) se convierte en 0, de modo que  $e^0$ , como cualquier número elevado a la potencia cero, es igual a 1. Eso hace que en los casos en que el índice se refiere a un solo vocablo, la fórmula de dividir la suma de uno más el logaritmo del número de usos en un autor, por el denominador resultante de uno más el logaritmo del número de autores, permanece invariable. En cambio, si el índice que verificamos consta de 3 vocablos, la fórmula  $e^{(s-1)}$  pasa a ser  $e^2=7,6$ , lo que nos está expresando la relación proporcional entre usar un vocablo aislado o usar una secuencia de tres vocablos, que sin duda es mucho más significativo. Podemos hacer una prueba de que esta proporción 7,6:1 no es exagerada verificando en el CORDE las diferencias cuantitativas entre usar un sustantivo o usar el sintagma que incluye al sustantivo, un especificador y un adyacente, bien un verbo y el sintagma nominal que le sigue en función de objeto directo. En el desenlace de *La tía fingida* encontramos la palabra, muy poco frecuente, 'pelaza', que el CORDE registra nueve veces; si ampliamos la búsqueda al sintagma 'la pelaza' (el más general), tenemos ya cinco casos, y si buscamos la se-



cuencia 'trabada la pelaza' no aparece ningún registro, ya que *La tía fingida* no viene registrada en ese repertorio. Añadiendo el texto del manuscrito Porras tendríamos un único registro y la razón sería de 9:1 desde 'pelaza' > 'la pelaza' > 'trabada la pelaza'. Otros ejemplos, también de TF, indican que en realidad expandiendo el núcleo de un sintagma la proporción suele ser superior incluso a 9:1, como lo demuestra otro ejemplo de comprobación: la secuencia 'con tanto ahínco', que es índice válido ya que sólo aparece en cuatro autores. El vocablo 'ahínco' lo registra el CORDE (11/6/2013) un total de 54 veces, de las cuales 35 corresponden a Cervantes; el sintagma 'tanto ahínco' se registra diecisiete veces (doce de ellas en Cervantes) y finalmente la secuencia entera quince veces (doce en Cervantes). Tenemos aquí un ejemplo estupendo de que ese vocablo es **característico** de Cervantes y que el sintagma 'con tanto ahínco' es de uso **frecuente** en su obra. El rastreo en otros autores concurrentes para la atribución de *La tía fingida* nos indica que no es **frecuente** en ningún autor y que es de uso repetido en Juan de Jáuregui (cuatro), Luis de Cabrera (cuatro) y San Juan Bautista de la Concepción (cinco) y meramente **ocasional** en Avellaneda y López de Úbeda, que sólo la usan una vez. No aparece ni en Espinel, ni en Quevedo ni en Salas Barbadillo. Si filtramos exactamente 'con tanto ahínco' aparece doce veces en Cervantes y una sola en López de Úbeda. Al ser una secuencia de tres vocablos, el exponente sería  $e^2$  y la fórmula que da el coeficiente de Cervantes se multiplicaría por  $1 + \log_{12}$  y se dividiría por  $1 + \log_4$ , es decir,  $3,3:2,2$ , lo que nos da aproximadamente 1,7, mientras que en el caso de López de Úbeda tendríamos  $1 + \log_1$  dividido por  $1 + \log_4$ , es decir, aproximadamente 0,6. La diferencia entre estos índices corregiría sustancialmente el modelo binario, 1/0, presencia/ausencia del rasgo, según el cual ambos autores coincidirían en el uso del mismo índice, a favor de una ponderación que, en vez de ser meramente cuantitativa (12:1), por medio de la aplicación de la fórmula quedaría reducida a una proporción aproximada de 3:1. Curiosamente una cifra muy cercana al valor del propio número  $e=2,718$ .

## CONCLUSIONES

Luis Astrana Marín (1944: 37-38), en un interesante estudio crítico-filológico sobre la atribución de *La tía fingida*, apuntaba que «tampoco hay razón para incluir como genuina, en tanto un documento coetáneo no lo demuestre rotundamente, la sabrosa novela de *La tía fingida*, y a pesar de las numerosas y respetables opiniones emitidas a favor de su autenticidad». A pesar de que Astrana admite «algunas analogías de lenguaje, modismos y estilo que ofrece *La tía* con ciertas producciones cervantinas» y habla de una «novela tan excelente, de un estilo tan semejante al suyo o que no desdice de él» (1944: 40) se resiste a atribuirle la obra «a falta de un documento incontrovertible». Es obvio que si se descubriera documentación de archivo resultaría innecesaria la propuesta crítica basada en modelos teóricos. Los resultados de nuestra investigación muestran, a partir de elementos objetivos y de un análisis matemático basado en el álgebra lineal, que *La tía fingida* es obra de Cervantes 'en el estado actual de la teoría'. La primera aproximación, meramente cuantitativa ofrece un ochenta por ciento de coincidencia con los usos cervantinos



y un margen de variabilidad entre el diez y el quince por ciento en el resto de los autores. La aplicación de la fórmula que pondera las variables de frecuencia de uso, número de autores y número de palabras por secuencia, eleva estos porcentajes hasta un 95 por ciento para Cervantes. Un debate sobre este problema de atribución tendría que partir de la propuesta de otro autor que además de cumplir con las características coincidentes entre la obra cervantina y el macrotexto de TF en sus dos versiones, cumpliera también con algunas características de TF que fuesen ajenas a la obra cervantina. En tanto no exista dicha propuesta, la obra debe considerarse de Cervantes y escrita en el período 1600-1605, como las primeras versiones de *Rinconete y Cortadillo* y de *El celoso extremeño*, insertas también en el manuscrito Porras de la Cámara.

### Bibliografía

- APRÁIZ, Julián (1906) *Juicio de La tía fingida*, Madrid, Imprenta de los sucesores de Hernando.
- ASTRANA MARÍN, Luis (1944) «Sobre *La tía fingida*», *Cervantinas y otros ensayos*, Madrid, Afrodisio Aguado, pp. 37-48.
- CERVANTES, Miguel (1864) *Obras*, B.A.E., Madrid, Rivadeneyra.
- (2010) *Novelas ejemplares*, Barcelona, Crítica (2ª ed.).
- CORDE: REAL ACADEMIA ESPAÑOLA, Banco de datos CORDE (Corpus Diacrónico del Español) en línea, <http://corpus.rae.es/cordenet.html>.
- FERNÁNDEZ-GUERRA Y ORBE, Aureliano (1864) *Noticia de un precioso códice de la Biblioteca Colombina: Algunos datos nuevos para ilustrar El Quijote*, Madrid, Rivadeneyra.
- ICAZA, Francisco de (1928) *De cómo y por qué La tía fingida no es de Cervantes*, Madrid, Editorial Voluntad.
- ISLA GARCÍA, Virginia (2010) «A vueltas con *La tía fingida*», en *Hos ego versiculos feci*, Madrid, Iberoamericana/Vervuert, pp. 75-101.
- LAY, David C. (2007), *Álgebra lineal y sus aplicaciones*, Madrid, Pearson Educación (3ª ed.).
- La tía fingida*, X - *Nobela de la tía fingida* en el Códice “Poesías y relaciones varias del siglo XVII” de la Biblioteca Colombina de Sevilla, sign. 56-4-34, pp. 77-88.
- LÓPEZ, Freddy (2011) «Donde se muestran algunos resultados de atribución de autor en torno a la obra cervantina», *Revista Colombiana de Estadística*, Junio 2011, vol. 34, nº 1, pp. 15 a 37.
- MADRIGAL, José Luis (2003) «De cómo y por qué *La tía fingida* es de Cervantes», *Artifara*, 2.



A. Rodríguez López-Vázquez, «La novela de *La tía fingida* atribuida a Cervantes»

POOLE, David (2004) *Álgebra lineal. Una introducción moderna*, México-Madrid, Thomson.

RODRÍGUEZ LÓPEZ-VÁZQUEZ, Alfredo (2010) «Una refutación de las atribuciones del *Lazarillo* a Alfonso de Valdés, Hurtado de Mendoza y Arce de Otálora: la hipótesis de Fray Juan de Pineda», *Lemir*, 14, pp. 313-334.

---- (2011) «Un tríptico presuntamente cervantino: *La Jerusalén*, *Los habladores* y el *Auto de la Virgen de Guadalupe*», *Revista di Filologia e Letterature Ispaniche*, XIV, pp. 57-76.

---- (2013) «Dos entremeses de Cervantes: *Doña Justina y Calahorra* y *La cárcel de Sevilla*», en *Artifara*, pp. 4-17.

VENABLES, William N. y Brian D. RIPLEY, (1997), *Modern Applied Statistics with S-PLUS*, New York, Springer Verlag (2<sup>nd</sup> edition).

