

CERVANTES: LOS VIAJES Y LOS DÍAS



Este libro se ha beneficiado de las ayudas a la investigación concedidas al Proyecto I+D+i *Novela corta del siglo XVII: estudio y edición (y II)* (FFI 2013-41264-P), dirigido por Rafael Bonilla Cerezo.

Ayntamiento de Castro del Río

Sujeto e institución literaria en la edad moderna
(FFI2014-54367-C2-1-R)

© Edición de: Pedro Ruiz Pérez

© De los textos: sus autores

© de la primera edición: Prosa Barroca y SIAL Ediciones, 2016

Bravo Murillo, 123 • 28020 Madrid (España)

Correo electrónico: info@gruposialpigmalion.es

Pedidos: 91 535 70 53

© de la colección Prosa Barroca: Proyecto I+D+i FFI 2013-41264-P www.prosabarroca.es

© diseño de la colección y de cubierta: José Ramón Trujillo y Rafael Bonilla Cerezo

Cubierta: *Jose Moreno Carbonero, El escrutinio de la Biblioteca de Don Quijote, 1925.*

La reproducción total o parcial de este libro (incluido su diseño), su alquiler, su incorporación a un sistema informático, su transmisión o transformación en cualquier forma o por cualquier medio, sea éste electrónico, mecánico, por fotocopia, por grabación u otros métodos, sin la autorización previa y por escrito de los titulares del copyright, vulnera derechos reservados.

ISBN-13: 978-84-15746-34-8

Depósito Legal: M-37614-2016

Hecho en España (Unión Europea)

CERVANTES: LOS VIAJES Y LOS DÍAS

EDICIÓN DE PEDRO RUIZ PÉREZ



SIAL Ediciones



www.prosabarroca.es

Dirección

Rafael Bonilla Cerezo (Universidad de Córdoba)

Coordinación Científica y Editorial

José Ramón Trujillo (Universidad Autónoma de Madrid)

Comité Científico

Julia Barella (Universidad de Alcalá)

Cristina Castillo (Universidad de Jaén)

Davide Conrieri (Scuola Normale Superiore de Pisa)

Angela Fabris (Universidad de Klagenfurt)

Abraham Madroñal (Consejo Superior de Investigaciones Científicas. CSIC)

Michel Moner (Universidad de Toulouse-Le Mirail)

Antonio Rey Hazas (Universidad Autónoma de Madrid)

Florencio Sevilla (Universidad Autónoma de Madrid)

Lia Schwartz (Universidad de Nueva York, CUNY)

Paolo Tanganelli (Università degli Studi di Ferrara)

Miguel Ángel Teijeiro (Universidad de Extremadura)

Secretaría

Begoña Rodríguez (Universidad Alfonso X el Sabio)

La colección «Prosa Barroca» se rige por un proceso de evaluación y revisión anónima realizada por dos especialistas de prestigio en el área (*peer-review*), uno de los cuales pertenece a su Comité Científico internacional. Todas las ediciones críticas y los trabajos científicos publicados en la colección han superado esta revisión por pares y siguen los criterios de estilo y las normas éticas establecidas en su constitución.

ÍNDICE

TRABAJOS Y DÍAS CERVANTINOS. A MODO DE PREÁMBULO	9
Pedro Ruiz Pérez	
PRELIMINARES PARA UNA BIOGRAFÍA CIENTÍFICA	19
Jorge García López	
«NO SOY BUENO PARA PALACIO». CERVANTES Y EL MECENAZGO LITERARIO.....	31
Patricia Marín Cepeda	
DOS NOTAS SOBRE LA ENEMISTAD LITERARIA ENTRE CERVANTES Y BERNARDO DE LA VEGA.....	43
Ignacio García Aguilar	
CERVANTES Y LOS POETAS (I): ANTE EL «CANTO DE CALÍOPE»	63
Pedro Ruiz Pérez	
CERVANTES Y EL CANON AMERICANO: EL «CANTO DE CALÍOPE», EL <i>VIAJE DEL PARNASO</i> Y EL «DISCURSO EN LOOR DE LA POESÍA».....	85
Adrián J. Sáez	
AVELLANEDA DESDE LA ESTILOMETRÍA.....	97
Javier Blasco	
DE POSES, PINTURAS, GESTOS Y ESTILOS: ICONOGRAFÍA CERVANTINA EN <i>LA GALATEA</i> Y <i>EL QUIJOTE</i>	113
Frederick A. de Armas	
MEMORIAS CERVANTINAS DEL CAUTIVERIO: EN TORNO AL ENCUENTRO CON JUDÍOS Y SUS RESONANCIAS BÍBLICAS	127
Ruth Fine	
EL «MISTERIO [...] ESCONDIDO» DE LAS <i>NOVELAS EJEMPLARES</i>	141
Georges Güntert	

FORMA Y SENTIDO DEL <i>QUIJOTE</i>	155
Alberto Blecua	
LOS DUQUES ARAGONESES Y LA «VARA DE MEDIR» DE CERVANTES LA CULPABILIZACIÓN DE SEGUNDO NIVEL EN LA <i>SEGUNDA PARTE DE DON QUIJOTE</i> : (<i>TRAMAS DEL QUIJOTE –V–</i>)	177
Pierre Darnis	
CERVANTES Y LOS FEMINISMOS	221
Anne J. Cruz	
BIBLIOGRAFÍA GENERAL	233
LOS AUTORES.....	253
RESÚMENES DE LOS ARTÍCULOS.....	257

AVELLANEDA DESDE LA ESTILOMETRÍA

JAVIER BLASCO

Universidad de Valladolid

Si hacemos caso a la literatura suscitada por el *Quijote* de Avellaneda en los últimos 20 años, la cuestión de la autoría se ha convertido, sin duda, en el tema sobresaliente de la crítica interesada en el apócrifo. La cuestionada identidad de la persona que se esconde tras el nombre de Alonso Fernández de Avellaneda en la improbable portada (Vindel, 1937) de los ejemplares del *Segundo tomo del ingenioso hidalgo don Quijote de la Mancha* que en 1614 sale de las prensas de Comellas en Barcelona (o, aunque más improbable, de las de Felipe Roberto, en Tarragona, si hacemos caso del pie de la irregular portada) ha puesto en pie un rosario de nombres (Jerónimo de Pasamonte, Liñán de Rianza, Suárez de Figueroa, Baltasar de Navarrete, Lupercio Leonardo de Argensola, José de Villaviciosa, Lope de Vega, Quevedo, Tirso de Molina, Castillo de Solórzano, Ruiz de Alarcón...), que ha venido a hacer, si no más rica, sí más abigarrada la larga lista de candidatos que desde el siglo XIX, siempre sin el éxito que otorga el consenso, venían desgranándose: Alfonso Lamberto, Jerónimo de Pasamonte, fray Alonso Fernández, fray Luis de Aliaga, Valladares de Valdelomar y un largo número del que le ahorro noticia al lector (Sánchez Portero, 2006).

No es mi intención aquí ocuparme del estado de la cuestión sobre el debate de la autoría del falso *Quijote*. Remito al apartado que, con inteligencia sobrada, maduro juicio y buena dosis de ironía, le dedica a este asunto mi admirado amigo Luis Gómez Canseco (2014), maestro de siglodoristas. Esto me ahorra tener que mediar en una fatigosa disputa en la que los egos pujan, como poco, tanto como los argumentos. Sólo diré que, en mi opinión, ninguna de las propuestas de autoría (ninguna de las más recientes, al menos) carece de base razonable sobre la que sustentarse, pero todas ellas (todas, incluida la mía a favor de Baltasar de Navarrete) ofrecen incontables flancos débiles, sin alcanzar superar ninguna de ellas la frontera de ese territorio gobernado por la ficción, más que por la certeza, que se corresponde con la geografía de la conjetura y del «podría ser».

Sigo pensando, como en 2005, cuando intervine en la disputada cuestión de la autoría del falso *Quijote* (Blasco: 2005b), que, mientras no contemos con un documento fiable que resuelva la contienda, el único material que puede proporcionar información objetiva y sólida es el propio texto del *Quijote*, y muy especialmente el material que conforma su discurso; esto es, el uso que el autor hace de la lengua: el idiolecto del apócrifo. Algunos pasos importantes se han dado en los últimos tiempos en los estudios de aquellos que se han interesado por el tema de la identificación de la persona que se esconde tras el nombre de Alonso Fernández de Avellaneda, y en tal camino se ha reconocido el carácter pionero –al menos, por cronología– de algunos trabajos de Madrigal (2005) y, sin modestia, también de alguno mío propio (Blasco, 2005b; Blasco y Ruiz Urbón, 2009), a los que pronto siguieron otros (entre otros muchos, Martín Jiménez, 2007; Rodríguez López-Vázquez, 2011).

Aunque de forma más intuitiva que científica, estos trabajos pretendían situarse en el contexto que trazan las siguientes palabras de Patrick Juola (la cursiva es mía):

Determining the author of a particular piece of text has raised methodological questions for centuries. Questions of authorship can be of interest not only to humanities scholars, but in a much more practical sense to politicians, journalists, and lawyers as in the examples above. Investigative journalism, combined with scientific (e.g., chemical) analysis of documents and simple close reading by experts has traditionally given good results. *But recent developments of improved statistical techniques in conjunction with the wider availability of computer-accessible corpora have made the automatic and objective inference of authorship a practical option* (2006: 235).

Frente a los métodos tradicionales, y a diferencia de lo que pretendía hacer la vieja estilística, determinadas técnicas estilométricas hoy nos permiten analizar con gran fiabilidad y objetividad determinados fenómenos del discurso sobre los que el autor tiene muy poco control y que, en consecuencia, son difícilmente objeto de una posible imitación¹. Sin entrar ahora en la especificación de cada una

¹ David L. Hoover y Shervin Hess (2009) han llamado la atención sobre algunas de las circunstancias que colocan los análisis cuantitativos entre interrogantes. Citan para ello unas interesantes palabras de Rudman (1997: 354) con prevenciones acordes con el sentido común: «If you are planning a study and cannot get the correct electronic texts, or you realize that control texts do not exist, do not do the study». Sin embargo, no hemos establecido todavía en los estudios sobre textos españoles que yo he tenido ocasión de conocer un protocolo de actuación de mínimas garantías en relación a las ediciones empleadas; al tratamiento del texto seguido por el analista; a la consideración de la posibilidad del error de copia, del OCR o de mecanografiado. En fin, «only if all of the analytic and control texts are valid can the results be trusted» (Juola, 2003b: 176). Y estas garantías mencionadas, tan puestas en razón, requieren, primero, disponer de un corpus fiable (ediciones autorizadas), elaborado a partir de textos cuyo proceso de regularización (si existe) asegure al menos una uniformidad gráfica, ortográfica y ortotipográfica (división de palabras, uso de negritas o de cursivas, etc.) que neutralice ciertos usos de escritura ajenos al autor, procedentes de la copia para la imprenta o de la persona que compuso la edición impresa. Además, habremos de estar seguros

de estas técnicas, sí que diré que todas ellas se basan en la aplicación al corpus analizado de ciertos algoritmos estadísticos, realizado por medios informáticos: esto es, sin que la subjetividad del analista intervenga en ningún caso en la selección de los fenómenos (palabras, expresiones, formulismos, etc.) objeto de análisis, quedando reducido el papel del investigador a la interpretación de los datos resultantes.

Los algoritmos estadísticos sobre los que trabajan los distintos programas informáticos aplicados a los estudios de atribución (Voyant, el paquete Stylo para R, CopyCatch, etc.) se basan siempre en la confrontación del texto objeto de nuestro análisis (texto *dubitado*) con otro u otros de autores conocidos (textos *indubitados*) sospechosos de pertenecer a la misma mano que el texto primero, con la idea de que aquel *texto indubitado* cuya distancia sea menor con nuestro *texto dubitado* tendrá más probabilidades que el resto de compartir paternidad con él. Para ello, medimos, por ejemplo, las palabras más frecuentes (100, 500, 1000 palabras más frecuentes), la distribución de dichas palabras en distintas porciones de texto, las coocurrencias de voces asociadas, la puntuación, etc.

Un buen ejemplo práctico de la efectividad de la estilometría aplicada a los estudios de atribución puede verse en un reciente trabajo de Nanette Rißler-Pipka (2016). A fin de mostrar la efectividad de la herramienta de análisis (el paquete Stylo de R), para su trabajo elige 31 textos narrativos (con la única excepción del *Viaje del Parnaso*) de cuatro autores del siglo XVII cercanos en el tiempo (Cervantes, Castillo Solórzano, Céspedes y Meneses, y Salas Barbadillo), con el resultado (p. 33) que muestra el siguiente dendrograma (a partir de la consideración de las 1500 palabras más frecuentes valoradas estadísticamente desde el algoritmo Eder's Delta²) (Cuadro 1):

Cuadro 1

Como se puede observar, las obras de un mismo autor aparecen siempre e inequívocamente asociadas. Rißler-Pipka realiza un nuevo ensayo (p. 34),

de la representatividad de los textos del corpus de modo que los textos *indubitados* pertenezcan a registros y géneros lo más cercanos posible a los del texto *dubitado*. Por otro lado, las herramientas de análisis computacional deberán haber mostrado «en laboratorio» su fiabilidad.

² Eder's Delta es una variante del algoritmo Delta de Burrows (2002), que se basa en la utilización de las frecuencias relativas de las palabras más comunes para establecer una propuesta de atribución de autoría. El algoritmo se basa en la media de las diferencias absolutas entre los índices de frecuencia de un conjunto de variables de palabras en *n* textos dados (*dubitados*) y los índices de frecuencia del mismo conjunto de variables de palabra en otro texto de referencia (*indubitado*). El texto *dubitado* con el menor valor de Delta tiene la probabilidad más alta (de todos los textos de la prueba) de mantener una relación genética con el texto *indubitado*. Delta ha sido sometido a pruebas exhaustivas por Hoover (2004a) y se está convirtiendo en una herramienta importante para estudios estilométricos. En otro estudio Hoover (2004b) observó alternativas o variantes de Delta y mostró que varias de estas variantes eran igual de eficaces que el original.

incluyendo ahora en su corpus de análisis el poema heroico *España defendida*, de Suárez de Figueroa, y el *Quijote* de Avellaneda (Cuadro 2):

Cuadro 2

Ni el género ni el tema de los textos parece fundamental en las asociaciones que establece Stylo, pero sí que resulta determinante la prosa y el verso. Todas las obras cervantinas en prosa se siguen asociando en una misma rama, en tanto que el *Viaje del Parnaso* se escapa de la «familia Cervantes» y tiende a asociarse con otros textos del corpus también en verso, de otros autores. El *Quijote* de Avellaneda, en cambio, se asocia inequívocamente con los dos *Quijotes* cervantinos. Y un caso curioso, pero dentro de esta misma lógica, lo ofrece *La Galatea*, obra en la que la abundancia del verso determina un resultado irregular.

Avanzando un paso más, Ribler-Pipka (p. 47) incorpora al corpus de referencia de su análisis dos textos nuevos de Quevedo (*El Buscón* y los *Sueños y Discurso*), las dos partes del *Guzmán de Alfarache*, el *Guzmán* apócrifo, *La Pícaro Justina*, el *Marcos de Obregón* de Espinel, la *Mosquea* de Villaviciosa y la *Vida* de Pasamonte. El resultado del análisis (5000 Most Frequent Words, Eder' Delta) es el siguiente (Cuadro 3):

Cuadro 3

La comparación de este cuadro con el anterior (junto a otras varias mediciones y distintas formas de representación gráfica de la distancia entre los textos, para cuyo examen remito al trabajo original de Ribler-Pipka) permite sacar algunas conclusiones que se hallan entre lo más valioso del medido y riguroso trabajo de la autora: tras constatar mediante un análisis impecable «que hay una semejanza entre el estilo de Cervantes y el de Avellaneda que no esperábamos» (p. 50), alcanza una conclusión que comparto totalmente: «Los experimentos digitales han mostrado que la diferencia no puede residir en la frecuencia del uso de las palabras. Así pues, ahora toca empezar a reflexionar sobre cómo se puede sacar provecho de métodos tan desiguales para llegar a tener nociones más claras y objetivas de lo que entendemos por el estilo literario de un autor» (pp. 50-51).

Posiblemente, la única objeción reseñable que puede ponerse al trabajo de Ribler-Pipka sea el no haber incluido entre las muchas variantes de análisis ensayadas (que afectan al corpus y al tamaño de la lista de palabras más frecuentes) los *grams* de palabra (casos de *verbatim* de 2, 3 o 4 elementos); y sobre todo el haber reducido lo que ella llama estilo a la frecuencia de palabras. Pero, desde luego, es el camino que ella señala el que hay que seguir transitando para llegar a definir con mejor precisión (como ella reclama en las palabras recién citadas) el idiolecto de un autor.

En esa línea, quiero situarme sobre sus huellas para ver si es posible dar algún paso más de aquellos que con tanta seguridad ella dejó establecidos. Pero antes de ponerme en camino me permito recordar alguna cuestión de carácter propedéutico sobre la naturaleza de las herramientas y de los análisis que con ellas podemos realizar. Para que el análisis estadístico y estilométrico ofrezca resultados con ciertas garantías de objetividad, hay que proceder con rigor en la constitución del corpus de análisis, en el tratamiento del mismo y, sobre todo, en la interpretación de los resultados. Enseguida me referiré a ello, pero antes enunciaré algunas reglas derivadas de la experiencia en el uso la herramienta y del sentido común.

1. La efectividad del análisis con Stylo depende en alto grado de la riqueza del corpus que respalde nuestros análisis. El hecho de que Stylo asocie inequívocamente el *texto dubitado* con un determinado *texto indubitado* no quiere decir, por sí mismo, que ambos tengan la misma paternidad. Quiere decir, sólo, que de todos los textos analizados ese texto indubitado es el más cercano. Pero siempre cabe la posibilidad de que el autor de nuestro *texto dubitado* no esté representado en el corpus de análisis. La inclusión de un autor nuevo puede hacer que el texto objeto de nuestro análisis pase a asociarse con el texto recientemente incorporado al corpus, si la distancia estadística de palabras más frecuentes es menor con este nuevo texto que con el primero.

2. Para que Stylo sea fiable es necesario un corpus de respaldo suficientemente representativo. Si el autor del *texto dubitado* se halla representado con otro texto en nuestro corpus, aunque se trate de un texto de contenido y género diferente, Stylo, con una probabilidad muy alta, asociará ambos textos.

3. Cuando se trata de textos de los siglos de oro hay que extremar las cautelas a la hora de estudiar el idiolecto de un autor y de interpretar los resultados estilométricos que las herramientas de análisis nos proporcionan: la puntuación (de enorme funcionalidad discriminadora de autoría) no es obra casi nunca del autor, sino del editor. Y lo mismo puede afirmarse de toda otra serie de cuestiones ortográficas (modernización ortográfica) que resultan fundamentales y que pueden «contaminar» el corpus. Y no se trata sólo de la intervención del editor moderno en el texto (cuya distorsión más o menos podría reducirse unificando la orto-tipografía de la totalidad del corpus), sino que también ha de contemplarse otra objeción (y ésta resulta insalvable): la intervención en el texto mismo de manos (como la del amanuense que hace la copia de imprenta o como la de los componedores del impreso) ajenas a la del autor. El laísmo o el leísmo, por ejemplo, raramente podrán servirnos para discriminar entre autores, porque tampoco podemos estar seguros nunca de que sea el autor y no el que compone la página para la imprenta, o en su caso el copista que ha intervenido en su transmisión, el que decida³.

³ No obstante, conviene considerar que, al trabajar con grandes listas de palabras (que pueden llegar hasta las 5000 más frecuentes), el significado estadístico de intervenciones ajenas al autor reduce el grado de incidencia, y este tipo de análisis (que en esencia es el que está en la base del paquete Stylo) sí que se ha mostrado rentable y eficaz.

4. Generalmente, se viene admitiendo que el léxico tiene una relevancia relativa, pues es lo más fácil de imitar. De hecho, parámetros como los de la riqueza léxica (que son el resultado de la aplicación de una contrastada fórmula matemática con varias opciones de planteamiento: con palabras de función sólo o con la totalidad del cuerpo léxico) son los que han demostrado menor funcionalidad y menor índice de acierto. Sin embargo, yo no estoy seguro de que los distintos análisis de que es susceptible el léxico resulten poco significativos: los análisis de Ribler-Pipka muestran muy claramente cómo el *Quijote* de Avellaneda, que inicialmente se asocia con los dos *Quijotes* del alcañino, cuando se eliminan del corpus de análisis los dos *Quijotes* cervantinos, lejos de asociarse con Cervantes, se asocia con el *Marcos Obregón* y con *La pícara Justina* (Cuadro 4):

Cuadro 4

Eso quiere decir que lo que asociaba a Avellaneda con Cervantes era la proximidad de contenido entre los tres textos; no otra cosa. Por ello, al eliminar los *Quijotes* de 1605 y 1615 el texto de Avellaneda deja de asociarse con otro texto de Cervantes y busca la asociación con otros más próximos, tales como el *Marcos Obregón* o *La pícara Justina*.

5. Los textos deben trabajarse (aunque con toda cautela de no alterar con ello el idiolecto) al menos en los siguientes aspectos: eliminación de preliminares (por ser en principio de mano ajena a la del autor), así como los nombres propios que preceden las intervenciones de los participantes, por ejemplo, en un diálogo; y, sobre todo, dado que Stylo se revela muy sensible a la diferencia entre prosa y verso (lo que ocurre en el análisis de Ribler-Pipka con el *Viaje del Parnaso* resulta definitivo), deberían eliminarse para el análisis todos los poemas⁴.

6. En cualquiera de los casos, Stylo (y el concurso de otras varias herramientas estilométricas, como las señaladas más arriba) servirán para ponernos (cuantitativamente) en la pista de aquellos textos con los que el que nos interesa guarda menores distancias. Sólo en segunda instancia, deberemos volver al texto para hacer otro tipo de valoraciones (cualitativas), en las que, desde luego, la ideología, la geografía, las lecturas, etc. dejarán un espacio importante a la conjetura hermenéutica. Pero actuar al revés, como se viene haciendo, es sacralizar la conjetura y, con ella, la subjetividad del investigador.

A partir de estos presupuestos, cuya fiabilidad ha sido suficientemente contrastada⁵, analizo a continuación el texto de Avellaneda confrontándolo con mi propio

⁴ Quizás deberían eliminarse también los epígrafes que presiden los distintos capítulos, dado que en muchos casos sabemos que son obra del impresor y no del autor, pero en una estadística que contempla listas de palabras más frecuentes tan copiosas la incidencia de tales paratextos es mínima.

⁵ Ni el contenido ni el género parecen ser determinantes en análisis basados en frecuencia de palabras con herramientas como Stylo. Esto afirman Hoover y Hess (2009: 473) y claramente lo prueban los dendrogramas de su estudio: «This analysis also shows that, even though nearly all of the 400

corpus (diferente al contemplado por Rißler-Pipka) e incorporando a mi análisis varios parámetros que puedan complementar lo alcanzado en el trabajo reseñado (por ejemplo, las coocurrencias de 2, 3 y, en su caso, 4 palabras), con un objetivo muy concreto: examinar desde la estilometría ciertas hipótesis críticas que han hecho fortuna en la bibliografía sobre el apócrifo. A saber: comprobar cuantitativamente si el idiolecto de las novelitas interpoladas en el falso *Quijote* («El rico desesperado» y «Los felices amantes») tienen, o no, una autoría diferente a la del resto del libro, como se ha sugerido por parte de notables críticos (Marín, 1981; Gómez Canseco, 2000).

Comenzaré por esta última cuestión. Separo el texto de las dos novelitas interpoladas en el *Quijote* de Avellaneda y creo un corpus (ACorpus) constituido por:

1. «El rico desesperado».
2. «Los felices amantes».
3. El *Quijote* apócrifo (sin el prólogo y sin las dos novelitas arriba mencionadas), dividido aproximadamente en dos ficheros de parecida extensión (I y II).
4. El «Prólogo» de Avellaneda.

El dendrograma (100MFW, Eder's Delta) resulta clarificador: asocia inequívocamente, por un lado, los dos archivos en que dividimos el *Quijote* apócrifo y, por otro, las dos novelitas interpoladas, en tanto que el «Prólogo», queda descolgado (Cuadro 5):

Cuadro 5

Y el mapeo estadístico no varía significativamente si aumentamos la lista de palabras más frecuentes a 1000. Pero con 2000 MFW y 4-grams el «Prólogo» se incorpora a la rama de las novelitas interpoladas y se asocia con la de «El rico desesperado» (Cuadro 6)

Cuadro 6

Si recurrimos a otro modo de mapear los resultados (*multidimensional scaling*), la distancia entre las dos novelas y el prólogo prácticamente se reduce a cero, en tanto que se magnifica la distancia entre estos textos y los dos en que hemos dividido el *Quijote* de Avellaneda (tras eliminar los poemas y extraer las novelitas interpoladas). Véase en el Cuadro 7 la superposición prácticamente absoluta de las dos novelitas interpoladas en el *Quijote*:

Cuadro 7

words [...] are content words, the different subject matter in the [...] texts does not prevent texts by the same author from clustering. Given the large number of texts, their varied subject matter and point of view, and the fact that some are fiction and some non-fiction, these are strong results».

Desde luego, todo parece confirmar la intuición de algunos críticos (Marín, 1981; Gómez Canseco, 2014) respecto a la posibilidad de que el prólogo y las novelas interpoladas pudieran ser de distinta mano a la que escribió la fábula del *Quijote* apócrifo. Pero es pronto todavía para concluir nada. Ninguno de estos resultados es realmente significativo, ya que Stylo trabaja asociando por pares las muestras disponibles en el corpus. Habrá que ver qué ocurre cuando aumentemos y cambiemos el contenido de nuestro corpus.

Procedemos, pues, a crear un nuevo corpus de análisis (Ccorpus) con los textos del «Prólogo» de Avellaneda a su *Quijote*, y los de «El rico desesperado» y «Los felices amantes» y otros textos de novela corta del momento (la idea es evitar al máximo las distorsiones que puedan derivarse de fórmulas narrativas ajenas a la novela corta). Y, así, junto a las dos novelitas objeto de estudio incluimos textos de Cervantes (varias *Novelas ejemplares*, con representación tanto de las llamadas «idealistas» como de las «realistas»), Cortés de Tolosa, Lope de Vega, Montalbán, Salas Barbadillo, Piña, Lugo y Dávila, Castillo Solórzano, Tirso de Molina, Zayas y Camerino. Vamos a contemplar las 1000 palabras más frecuentes. El resultado (Cuadro 8) no deja de tener interés.

Cuadro 8

Subrayo, a la vista de la Cuadro 8, algunos resultados dignos de reseña:

1. Por lo general, cada texto ha sido asociado inequívocamente con otro texto del mismo autor que el primero. Así ocurre con Cortés de Tolosa, Lope, Salas Barbadillo, Zayas, Camerino. Las excepciones o corresponden a textos sin pareja o se trata de casos (*Los sueños de Quevedo* y «La tía fingida», muy cerca de los Cervantes) merecedores de una reflexión aparte, que ahora me reservo para no complicar el objetivo principal del análisis.

2. El programa claramente ha separado el texto del prólogo de Avellaneda, asociándolo con el de las *Ejemplares* de Cervantes, del resto de textos narrativos. Esto parece indicar que el algoritmo estadístico empleado (Delta) es sensible a la extensión y, quizás, al género de los textos.

3. Los textos de las dos novelitas que nos ocupan los ha situado en una rama en la que se encuentran también los textos de Lope, Montalbán, Piña, Lugo y Dávila, Salas Barbadillo, Zayas, Camerino y dos más que el algoritmo estadístico asocia de manera muy particular: Castillo Solórzano y, sobre todo, Tirso.

Si aumentamos a 5000 la lista de palabras más frecuente, las variaciones que se producen respecto al dendrograma anterior no resultan sustanciales, pero sí ofrecen algún matiz interesante (Cuadro 9):

Cuadro 9

Interesante es lo que ocurre con «La tía fingida», aunque, insisto, no es esta obra la que ahora nos importa. Lo relevante, en relación con los textos de Avellaneda estudiados, es que las asociaciones que mostraba el cuadro anterior (con 1000 palabras más frecuentes) se mantienen bastante estables.

A partir de la información que proporcionan los dendrogramas anteriores, no afirmaré (¡Dios no me deje caer en esa tentación!) que fue Tirso el autor de esas novelitas interpoladas, que siempre se pensó que eran ajenas al discurso principal de la fábula de Avellaneda. Y no lo haré, primero, porque la proximidad del «tordesillesco» Castillo de Solórzano obliga a la prudencia y, segundo, porque el análisis cuantitativo que subyace a los cuadros «apunta tendencias» (y lo hace con una fiabilidad altísima), pero nunca es absolutamente concluyente: no prueba que Tirso sea el autor de «El rico desesperado» y «Los felices amantes»; lo que prueba es que, de todos los autores contemplados, Tirso es el más próximo. Y, así, en el caso de que Tirso no fuere el autor, y un día subiéramos a ese mismo corpus una obra del autor de ambas novelitas, ellas se asociarían con esta nueva obra.

Lo que sí que podemos concluir de modo casi definitivo es que una sola y misma mano escribió los dos relatos objeto de nuestro análisis. Con el fin de afinar un poco más, procedo a continuación a preparar un nuevo corpus (Bcorpus), añadiendo sobre la base de Acorpus los textos íntegros de Avellaneda (es decir, el texto completo: *Avellaneda* + el mismo texto, sin la novelitas interpoladas y dividido en dos bloques con número similar de palabras: *Avellaneda_I* y *Avellaneda_II*) y aquellos otros textos (sólo los más relevantes) que la crítica ha puesto en relación con la autoría del *Quijote* apócrifo: el *Guzmán de Alfarache*, de Mateo Alemán; el *Guzmán* apócrifo; Castillo Solórzano, con las *Aventuras del Bachiller Trapaza*; Lope de Vega, con *Los pastores de Belén*; Vicente Espinel, con el *Marcos Obregón*; Baltasar de Navarrete, con *La pícaro Justina*; Suárez de Figueroa, con *El pasajero*; Tirso de Molina, con *Los Cigarrales de Toledo*; José de Villaviciosa, con *La Mosquea*. Los resultados no dejan de tener gran interés (Cuadro 10):

Cuadro 10

A vista de lo que se refleja en el dendrograma resultante, pueden hacerse dos anotaciones que resultan interesantes a la vista de lo que la crítica ha venido escribiendo:

1. El texto de Avellaneda, en su forma completa y en su forma expurgada y fragmentada, viene a situarse en solitario (y esto es importante) en una rama del árbol diferente al resto de los textos. Quiere ello decir, en principio, que el idiolecto de Avellaneda (si consideramos exclusivamente las 500 palabras más frecuentes) no se corresponde con el de ninguno de los autores de la otra rama (Lope, Castillo Solórzano, Tirso, Baltasar Navarrete, Pasamonte, Alemán, Espinel y Suárez de Figueroa).

2. A diferencia de lo dicho en el punto anterior, las novelitas interpoladas («Los felices amantes», «El rico desesperado») reflejan cierta proximidad al idiolecto de Tirso y de Castillo Solórzano.

Si aumentamos la cifra del listado de palabras más frecuentes que le pedimos a Stylo que considere (1000 ó 2000 palabras), las cosas no varían sustancialmente (Cuadros 11 y 12):

Cuadro 11

Cuadro 12

No obstante, el lector recordará que mi análisis comenzaba su andadura a partir de las conclusiones del trabajo de Nanette Ribler-Pipka, que trabajaba con un corpus en el que Cervantes jugaba un papel importante, en tanto que mi análisis ha dejado fuera del corpus a Cervantes. Procedo por ello a enmendar tal ausencia, creando un nuevo corpus (Ecorpus = Bcorpus + el *Quijote* 1605).

Pero, como puede observarse (Cuadro 13), nada cambia sustancialmente con la incorporación al análisis del *Quijote* cervantino:

Cuadro 13

Conclusiones

¿Es posible con estos datos llegar a alguna conclusión? Muy provisionalmente. Me explico. El análisis de las palabras más frecuentes (en porcentajes que oscilan en listados de 1000 a 5000 palabras) parece indicar que:

1. Tal y como siempre sospeché la crítica, desde la intuición formulada hace varias décadas por Nicolás Marín, las dos novelitas que incorpora el *Quijote* de Avellaneda no parecen salidas de la misma mano que la globalidad del texto del apócrifo.

2. Las dos novelas, «Los felices amantes» y «El rico desesperado», sí que proceden de la misma mano.

3. Enfrentando el texto de las novelitas con la práctica totalidad de autores que tiene un lugar en el panorama de la novela corta de las primeras décadas del siglo XVII, la escritura de Tirso y Castillo Solórzano es la que resulta más próxima⁶.

4. El idiolecto de Avellaneda (cuando se valora exclusivamente las listas de palabras más frecuentes) se revela muy alejado del de todos aquellos autores que se han barajado con más fortuna crítica.

⁶ En este punto sería oportuno utilizar el t-test de que Burrows (1992) se sirvió para distinguir entre dos posibles autores (James o Austen) en un anónimo, mediante la realización de la prueba t (t-test) a partir de los vocabularios combinados de grupos de textos de James y Austen, para identificar conjuntos de palabras que permiten diferenciar con claridad los dos autores.

Pero, he de insistir en ello, tales conclusiones sólo pueden tener un valor provisional. Joula, en 2006, manifestaba lógicas dudas respecto a la aplicación a la lingüística de algoritmos que ciertamente se había demostrado que eran valiosos matemática y estadísticamente⁷. Quizás hoy tales dudas están menos justificadas, pero sigue siendo verdad la necesidad de tener en cuenta en el análisis del idiolecto de un autor, además de las palabras más frecuentes, otros parámetros. Stylo se ha revelado como un paquete de análisis extremadamente útil y fiable para medir la distancia que media entre dos textos a partir de las palabras más frecuentes en cada uno de los textos. La posibilidad que ofrece de manejar amplísimas listas de frecuencia, así como la probada rentabilidad del algoritmo Delta o Eder's Delta desde el que se tratan los datos, garantizan unos resultados dignos de consideración. Pero no resulta fácil aceptar que el idiolecto de un texto pueda concretarse exclusivamente, y en última instancia, en la frecuencia léxica. La siguiente consideración de Joula (2006: 172) refleja puntualmente las razones de mi cautela respecto a unos análisis estadísticos que se basan exclusivamente en el léxico: «Finally, methods based on simple lexical statistics tended to perform substantially worse than methods based on N-grams or similar measures of syntax in conjunction with lexical statistics. We continue to examine the detailed results in an effort to identify other characteristics of good solutions».

Varios especialistas, con relevantes estudios de atribución (Burrows, 1989; Kesselj and Cercone, 2004; Grieve: 2007), han insistido en la conveniencia de combinar varios parámetros para otorgar cierta fiabilidad a un análisis cuantitativo. Y en ese sentido ciertas herramientas (como CopyCath, Voyant, WordSmith, y JGAAP) nos permiten trabajar variables diferentes a la frecuencia de las palabras y cuantificarlas. Así con JGAAP he procedido a hallar, en los textos que conforman Bcorpus, el *coeficiente de variación*⁸ de las siguientes variables: Primera palabra de frase, NGrams de palabra (n : 2), vocal inicial de palabra, NGrams de caracteres (n : 2), vocales M---N (m : 2, n : 3) y letras M--N (m : 1, n : 2). Los resultados obtenidos muestran, como era de esperar, que el primer fragmento del *Quijote* apócrifo (Avellaneda_I⁹) es el que ofrece menor distancia respecto al texto

⁷ Y añadía: «Unfortunately, another apparent result is that the high-performing algorithms appear to be mathematically and statistically (although not necessarily linguistically) sophisticated» (Joula, 2006: 172).

⁸ Analiza [$\text{CoV} = (\sigma/\mu) \cdot 100$] los textos objeto de análisis cuantificando la desviación estándar de cada uno de ellos (Lope, Alemán, Espinel, Tirso, Pasamonte, etc.) respecto a otro dado (Avellaneda) que se toma como referencia. El idiolecto de aquel cuya desviación sea menor ofrecerá una probabilidad mayor de autoría compartida.

⁹ Estas son las cifras del coeficiente de variación de *Avellaneda I*: First Word In Sentence: 23.600503469647744; Sorted Word NGram (n : 2): 24.326448882980156; Word NGrams (n : 2): 26.836941341095393; Vowel-initial Words: 16.433118682191818; Character NGrams (n : 2): 12.537061416895911; Vowel M--N letter Words (m : 2, n : 3): 15.194197310104107; M--N letter Words (m : 1, n : 2) : 13.22696747940104.

completo de Avellaneda en todos los parámetros objeto de análisis, seguido por Tirso (Cuadro 14):

Cuadro 14

Desde luego, estos resultados nos obligan, si no a corregir, sí a reconsiderar alguna de las conclusiones a las que se llegaba a partir de Stylo (especialmente las relativas al punto cuarto) por el relieve que ahora los análisis de varios parámetros otorgan al autor de los *Cigarrales de Toledo*: de todos los candidatos analizados (Aleman, Martí, Navarrete, Suárez de Figueroa, Pasamonte, Espinel, Castillo Solórzano), Tirso de Molina ofrecen el más bajo coeficiente de variación respecto a Avellaneda.

En principio, si Tirso fuese el autor del falso *Quijote*, resultaría curioso, sin duda, lo que en *El amor médico* afirma de la escritura sin firma que se responsabilice de ella: «Carta sin firma es libelo / que contra sí mismo hace / quien no osa poner su nombre / por confessar que es infame [...] / Lo que no firmaron plumas / firme el acero, y no manchen / espejos de honor honestos / cartas que sin firma salen».

En cualquier caso, tendremos que tener en cuenta que la proximidad de Tirso a Avellaneda que refleja el cuadro anterior puede tener su lógica si Tirso es el autor (como parecía apuntar Stylo) de las dos novelitas interpoladas que forman parte del texto tomado como referencia en nuestro análisis del coeficiente de variación. Dichas novelas ocupan una porción del texto del falso *Quijote* que resulta suficiente para que los nuevos análisis reflejen una notable proximidad con el apócrifo. La mano de Tirso está en una parte del falso *Quijote*, y ello es lo que reflejan estos análisis¹⁰. De todos modos, conviene no poner el foco en un autor precipitadamente¹¹ y tener en cuenta prevenciones como las que David L. Hoover

¹⁰ Sin entrar en el fondo de la cuestión, pero recordando que la atribución del texto de Avellaneda a Tirso es antigua, aquí cabe una consideración no «cuantitativa», aunque sólo sea para recordar que Tirso es gran amigo de Lope (y apasionado defensor de su fórmula dramática), como siempre se ha sospechado de Avellaneda; que conoce bien todo el trayecto que don Quijote sigue de Zaragoza a Madrid, así como la geografía urbana de Toledo que Avellaneda muestra conocer; que su ideología y sus «teologías» se ajustan bastante bien a lo que refleja el apócrifo; que en 1614 se ve obligado a retirarse al monasterio de Estercuel por problemas con las autoridades religiosas que bien podría explicar la alusión de Cervantes en el prólogo del *Quijote* de 1615, cuando refiriéndose al falsario, se dirige al lector con estas palabras: «Paréceme que me dices que ando muy limitado y que me contengo mucho en los términos de mi modestia, sabiendo que no se ha de añadir aflicción al afligido y que la que debe de tener este señor sin duda es grande, pues no osa parecer a campo abierto y al cielo claro, encubriendo su nombre, fingiendo su patria, como si hubiera hecho alguna traición de lesa majestad».

¹¹ Basta comprobar cómo los resultados de mi trabajo difieren de los de Ribler-Pipka, a pesar de trabajar ambos con la misma herramienta, por la sencilla razón de que el corpus de análisis suyo y mío difieren. Elocuente resulta lo que sucede en el análisis de Ribler-Pipka con el *Viaje del Parnaso*,

y Shervin Hess (2009) han manifestado acerca de análisis computacionales como el que aquí he realizado. Pero un análisis con Stylo (o con JGAAP), al producir unos resultados (más o menos nítidos) sin intervención alguna de la subjetividad del investigador, permite orientar futuros análisis cualitativos de atribución en una determinada dirección, reduciendo notablemente el abanico de candidaturas. A la vista de los resultados que reflejan los cuadros 11, 12 y 13, la autoría del *Quijote* de Avellaneda habrá que buscarla, en el futuro, en otros territorios diferentes a aquellos que la crítica especializada ha explorado hasta el presente; pero, a la vez, los cuadros 8 y 9 muestran de manera bastante clara que Tirso de Molina anduvo muy cerca del apócrifo y de modo altamente probable fue el autor de «Los felices amantes» y «El rico desesperado», las dos novelitas interpoladas en el texto de Avellaneda¹².

que inequívocamente se asocia con el resto de obras de Cervantes sólo hasta que se añade al corpus una obra en verso como la de Suárez de Figueroa (*España defendida*). Véase Cuadro 2.

¹² Menos segura es la presencia de su mano en el prólogo.



Esta primera edición de
Dedicado a Cervantes de Aldo Ruffinatto,
acabose de imprimir el 1 de junio 2015,
día de santa Laura, inspiradora de poetas.

LAUS DEO HONOR ET GLORIA

