

## RESEÑAS

**Miguel Garci-Gómez. *Tres autores en 'La Celestina': Aplicación de la informática a los estudios literarios*. Granada: Colección Romania. 1993. 267 pp.**

El autor de este libro estudia el léxico de *Celestina*, trazando la distribución de frecuencias con ordenador para llegar a la conclusión de que tres partes de la obra fueron compuestas por distintos autores. Los estudiosos de *Celestina* encontrarán aquí un repaso completo del estado de la cuestión de múltiple autoría, más un gran acopio de datos relacionados con el vocabulario de la obra que serán de enorme utilidad para discutir temas y estilo, pero el libro dista de poner fin a dicha cuestión.

Garci-Gómez muestra profundos conocimientos de la obra en sí, además de gran pericia filológica. En dos capítulos examina vocablos discretos relacionados a temas, organizando su discusión y análisis de vocablos según categorías semánticas de la contienda, la caza, carnes, la cocina, la bestialidad, la religión, la brujería, la desconfianza, el servicio, el parentesco, impresiones sensoriales, la exclamación y la risa. Otros cuatro capítulos tratan los apartes, grupos morfosintácticos, diminutivos, pronombres, adverbios y frases, formas verbales y formas ortográficas. El autor sabe bien qué tipo de material se presta a la búsqueda computacional y presenta los datos en cuadros muy claros.

Si por una parte el libro lleva una nutrida bibliografía sobre *Celestina*, por otra parte no hay ninguna sobre el uso del ordenador para resolver casos de autoría disputada, una disciplina que cuenta con una respetable tradición. El ya clásico estudio de Mosteller y Wallace sobre los *Federalist Papers* (1964)<sup>1</sup> y numerosas publicaciones sobre los apócrifos shakespearianos y textos bíblicos han proporcionado modelos del uso de la informática para casos de

---

<sup>1</sup> Frederick Mosteller and David L. Wallace, *Inference and Disputed Authorship: "The Federalist"*, Reading, Massachusetts: Addison-Wesley, 1964.

autoría disputada. El presentar este estudio de *Celestina* como "logoscopia" sin citar modelos anteriores ni atenerse a principios comúnmente aceptados en el campo de la atribución autorial hace difícil aceptar las conclusiones.

El autor divide el texto de *Celestina* en tres partes para llevara cabo su análisis: AUTO (Acto I), COMEDIA (Actos II-XVI) y TRATADO (material interpolado). El solo hecho de considerar *Celestina* para el análisis textual como una concatenación de tres obras diferentes, además de representar una aceptación *a priori* de la presencia de tres autores, es problemático. La única prueba que conozco que utiliza textos concatenados es QSUM, pero sólo para estudiar hábitos lingüísticos, como el porcentaje de sustantivos o de palabras de cierta longitud.<sup>2</sup> También cabe pensar que si un autor ha reelaborado y continuado un texto comenzado por otro, es probable que la parte original no nos haya llegado "pura" y que lleve señales estilísticas de su intervención. Como reconoce Garci-Gómez en su introducción, se trata de textos "enrevesados" (26). Por eso mismo es mucho más complicado que comparar dos dramas para determinar si el estilo se acerca más a Shakespeare o a Marlowe.

El libro presenta los datos de una sola operación analítica, los que produce *WordCruncher* para trazar la distribución de frecuencias. Los datos están organizados en cuadros, con columnas que corresponden a: (1) casos (número absoluto de ocurrencias), (2) porcentaje real del texto que constituye el número absoluto, (3) porcentaje esperado (de estar distribuidos de modo aleatorio en textos de extensión igual) y (4) la diferencia entre el porcentaje real y el esperado. Como el uso de esta función para determinar autoría representa una metodología nueva, conviene, antes de presentar los datos, asesorar la validez del experimento utilizado para estos fines. La distribución de frecuencias de *WordCruncher* está basada en la premisa de que tal distribución *en una obra* (no de una serie de obras concatenadas) será suficientemente homogénea para ser asimilada a una distribución aleatoria. Se emplea normalmente como una especie de mapa para localizar la concentración de un determinado elemento en distintas partes de un texto. Se puede comparar el patrón de la repartición con el de otro autor o texto, como ha hecho Muller con

---

<sup>2</sup> David I. Holmes. "Cumulative Sum Charts for Authorship Attribution," *ACH-ALLC Conference Abstracts*, Washington, DC: Georgetown University, 1993.

textos de Racine y Corneille,<sup>3</sup> siempre y cuando se emplea alguna prueba de significación, como la de  $\chi^2$ , que mide homogeneidad.

El autor no explica por qué decidió trabajar en dos capítulos con vocablos de contenido temático, que suelen ser muy variables para constituir un discriminante estilístico, en vez de seguir los modelos de atribución basados en el comportamiento de vocablos raros (Udne Yule<sup>4</sup>), funcionales (A. Q. Morton<sup>5</sup>) o de alta frecuencia que cumplen con una razón diferencial designada (Warren B. Austin<sup>6</sup>).

Otro problema es que la repartición de vocabulario puede ser muy irregular en un solo autor debido a causas temáticas y estilísticas (según señala una y otra vez Charles Muller en *Principes et Méthodes de la Statistique Lexicale*), sin que sea indicio de autoría distinta. Como la probabilidad de ocurrencia de un vocablo crece con la extensión del texto siguiendo una regla binomial, como ha demostrado Muller, se debe usar muestras de longitud igual para las comparaciones de vocabulario. Generalmente los estudios de atribución se ocupan de atributos que representen hábitos lingüísticos no afectados por temas específicos en el contenido del texto. En este sentido, los datos de Garci-Gómez sobre grupos morfosintácticos (prefijos, diminutivos, pronombres, etc.), formas verbales y formas ortográficas pueden tener validez, pero únicamente si se formula una hipótesis clara y se comprueba la significación estadística mediante alguna norma—  $\chi^2$ , desviación z, razón diferencial, índice de dispersión, cusum, etc.. Sin pruebas, los cuadros son interesantes pero no hay modo de juzgar si las diferencias rebasan las que normalmente se puedan esperar tratándose de un solo autor.

---

<sup>3</sup> Charles Muller, *Principes et Méthodes de la Statistique Lexicale*, Paris: Classiques Hachette, 1977.

<sup>4</sup> G. Udny Yule, *The Statistical Study of Literary Vocabulary*, Cambridge: Cambridge UP, 1944.

<sup>5</sup> Andrew Q. Morton, *Literary Detection*, New York: Charles Scribner's Sons, 1978.

<sup>6</sup> Warren B. Austin, "The Authorship of Certain Renaissance English Pamphlets," *Proceedings: Computer Applications to Problems in the Humanities*, ed. Frederick M. Burelbach, Jr., Brockport, NY: SUNY College, 1970, 93-99.

Establecer múltiple autoría exige también más de un tipo de prueba. En los estudios de vocabulario hay muchas posibilidades, tales como la razón de vocablos a palabras (TTR), hapax legomena, el crecimiento incremental de vocabulario (posible desviación de la norma teórica binomial de Muller) y medidas de vocabulario solapado hechas sobre muestras aleatorias iguales de cada parte. También se echan de menos información sobre la extensión en número de palabras de cada parte (sólo se señalan porcentajes), una explicación del uso de frecuencias inferiores (como tres o de cuatro ocurrencias [185 y 186]); la enunciación de una hipótesis nula y una explicación de la metodología empleada.

Garci-Gómez considera que las diferencias que se desprenden de los cuadros de repartición son "de suficiente peso como para establecer la participación de un tercer autor en *La Celestina*" (192), pero para alegar autoría múltiple, hay que establecer significación estadística. Los datos aportados permiten afirmar solamente que el léxico varía entre las tres divisiones, como varía también en una obra escrita por un solo autor.

Ciertos resultados parecen dignos de notar, como los que tienen que ver con variantes ortográficas y morfosintácticas, porque éstas, a diferencia del vocabulario, son hábitos que no están sujetos a una relación binomial a la longitud del texto. La contundente afirmación de que los que "sostengan que los numerosos datos y la diferencias aquí analizadas no son de suficiente peso como para establecer la participación de un tercer autor en *La Celestina* serán los prefieran seguir refugiados en el presentimiento, en la corazonada, en el conviene, luego es" (192) no está justificada. Los que confían en los ordenadores y en las estadísticas tampoco se convencerán. La enorme bibliografía en torno a los estudios de autoría disputada aconseja en contra de afirmaciones tan tajantes y a favor de cautela, ya que hasta tasas tan altas como el 99% de significación estadística aún dejan algún resquicio a dudas.

En fin, la presentación de sugestivos datos léxicos y su organización en reveladoras categorías semánticas representan una labor digna de encomio, pero la evidencia aportada por este libro ha enriquecido la polémica--no la ha resuelto.

Estelle Irizarry

Georgetown University