

LINGÜÍSTICA FORENSE Y CRÍTICA TEXTUAL. EL CASO AYALA–CERVANTES

TEXTUAL CRITICISM AND FORENSIC LINGUISTICS. THE AYALA-CERVANTES FILES

José Manuel FRADEJAS RUEDA

Universidad de Valladolid

josemanuel.fradejas@uva.es

Resumen: En este artículo se demostrará que la edición del *Libro de la caza de las aves* de Pero López de Ayala, publicado por la Biblioteca Virtual Miguel de Cervantes, no es una digitalización de la edición de José Gutiérrez de la Vega, como aducen, sino que se trata de una copia literal de la versión modernizada que José Fradejas Lebrero publicó en 1969. Además, se mostrará que el texto colgado por la Virtual Cervantes ha sido publicado, en papel y en formato e-book, por una editorial comercial y que no se trata de un derivado directo de la edición de Gutiérrez de la Vega, como sus editores indican, sino de mera copia exacta de la de la BVMC y, por ende, de la de JFL. Para llegar a estas conclusiones los textos se han analizado con software de uso en lingüística forense (colación automatizada, comprobadores de plagio y programas de concordancia), para lo que primero se expondrá qué es la lingüística forense y se pondrá en relación con la posibilidad de que los métodos de la crítica textual sean de valor para la lingüística forense.

Abstract: In this paper will be demonstrated that the edition of Pero López de Ayala's *Libro de la caza de las aves* uploaded by Biblioteca Virtual Miguel de Cervantes is not a derivative of José Gutiérrez de la Vega's 1879 edition, as they claim, but a *verbatim* copy of José Fradejas Lebrero's version published in 1969. Moreover, the uploaded text was published by a commercial publisher and is being sold as a derivative of Gutiérrez de la Vegas edition by a number of on-line resellers as printed book and as e-book. To prove it, all texts had been analysed with software used in forensic linguistics (automated collation, plagiarism checkers and concordancing software), but firstly it will explained what Forensic Linguistics is and later on will be discussed the possibility that the methods used in textual criticism are of value for Forensic Linguistics.

Palabras clave: Crítica textual. Colación automática. Lingüística forense. Plagio.

Key Words: Textual criticism. Automatic collation. Forensic linguistics. Plagiarism.

1. INTRODUCCIÓN

1.1 LA LINGÜÍSTICA FORENSE

La *lingüística forense* (LF) es una disciplina muy reciente y se encuadra dentro de la lingüística aplicada. La primera vez que aparecen los términos *lingüística* y *forense* juntos es en 1968, en el análisis que Jan Svartvik (1968) realizó sobre las confesiones que Timothy John Evans, un británico que fue sentenciado a muerte por el supuesto asesinato de su esposa embarazada en 1949, hizo ante la policía¹. Pero es hasta la década de 1990 cuando se fija como una disciplina académica, se fundan la International Association of Forensic Linguists (en 1992) y la International Association for Forensic Phonetics and Acoustics (en 1991) y se crea la revista *Forensic Linguistics: The International Journal of Speech, Language and the Law* (desde 1994)².

Esta disciplina, según la AIFL, se define como la interfaz entre Lengua y Derecho (Jiménez Bernal, Reigosa Riveiros y Garayzábal Heinze, 2014: 29; Turell, 2005: 13), lo cual no es no es muy clarificador ya que no queda claro el significado de *interfaz* (*interface*)³. Robert Leonard (2005: 65) lo explica de manera sencilla y con algunos ejemplos claros:

Forensic linguistics respond to legal questions that involve language, for example: Who wrote a ransom note? Who called in a bomb threat? What is the meaning of a phrase in a contract? If a person is searched by the police without a warrant, just what constitutes "consenting" to that search? If an interrogator indirectly makes a promise, does that invalidate a confession? Is "Mc" part of the McDonalds trade name or just part of English (and can I name my restaurant McHamburgers)? If a woman says, "I'm going to kill my rotten boyfriend," has she threatened his life? Could a fourth-grade dropout actually have written a confession with the phrase, "he approached the vehicle and I raised my weapon?"

Así, suele hablarse de una definición ancha de lingüística forense, que es la mantenida por la AIFL, que cubre todas las áreas en las que coinciden la lengua y el derecho y otra restrictiva que se refiere a la utilización de pruebas lingüísticas en los juicios (Blackwell, 2012: 4; Turell, 2005: 13). Esto ha llevado a Cotterill (2012), a plantear que la lingüística forense tiene dos ámbitos de

1 Según indican Coulthard y Johnson (2007: 5) el término *forense* apareció junto al nombre de una lengua casi 20 años antes, en el título del libro de Philbrick (1949).

2 El título inicial fue *Forensic Linguistics* pero en 2003 se cambió a *Forensic Linguistics: The International Journal of Speech, Language and the Law* (<http://www.equinoxpub.com/journals/index.php/IJSL>).

3 Téngase en cuenta que el *Diccionario de la Real Academia Española* solo considera la voz *interfaz* como un término informático que define como la 'conexión física y funcional entre dos aparatos o sistemas independientes', pero en inglés *interface* también puede significar, como sustantivo, 'a point where two systems, subjects, organizations, etc. meet and interact'.

atención: 1) lingüística forense descriptiva y 2) lingüística forense de investigación. La primera se centra en la “language of the legal process, from arrest to trial and beyond” mientras que la segunda “involves the analysis of texts as part of criminal investigations”, lo que viene a coincidir con la división de Coulthard y Johnson (2007) entre la lengua del proceso legal y la lengua como prueba. Es un amplio campo de trabajo en el que siempre los especialistas se enfrentan con un problema de análisis de textos, ya sean escritos ya sean orales (fonética forense), con implicaciones legales o vistas desde el campo del derecho.

Entre los muchos ámbitos en los que se aplica la lingüística forense se hallan los problemas de atribución de autoría y los de copia (plagio). Pero definir qué es plagio es complicado. Según el DRAE, plagiar es ‘copiar en lo sustancial obras ajenas, dándolas como propias’, pero desde el ámbito legal no hay definición alguna. Es más, la última ley de propiedad intelectual (Ley 21/2014, de 4 de noviembre; BOE 5.11.14) no menciona el plagio, aunque sí aparece en el artículo 270 del Código Penal vigente. Lo cierto es que en esto nos movemos un mundo nebuloso; hay tribunales que estiman que la noción de plagio es un juicio de valor, es decir, es algo subjetivo. Sin embargo, en la sentencia del Tribunal Supremo (n.º 1125/2003 26.11.2003) se establece que “el concepto de plagio ha de referirse a las coincidencias estructurales básicas y fundamentales y no a las accesorias, añadidas, superpuestas o modificaciones no trascendentales”.

Cómo demostrar que hay plagio (copia) puede ser labor de la lingüística forense y está íntimamente unida a los problemas de atribución de autoría (Pecorari, 2012), y para ello ha de aplicar todos aquellos métodos que la lingüística pueda utilizar para demostrarlo. Las técnicas más usuales son los análisis estadísticos como la estilometría (Holmes, 1998), la lingüística de corpus (O’Keeffe, McCarthy y Carter, 2007: 20) o cualesquiera otras que permiten medir la probabilidad (Olsson, 2004: 18) de que un texto de un autor u otro, sea una copia o no ya que los resultados que ofrecen “benefits from a scientific presentation in court. This includes the use of graphs and charts, tables and statistics” (Cotterill, 2012).

2. EL CASO AYALA-CERVANTES

2.1 ANTECEDENTES

A mediados de los años de 1990 empiezan a proliferar las llamadas bibliotecas electrónicas, digitales o virtuales. Por lo general, se suelen considerar estos tres nombres como sinónimos ya que su objetivo, en definitiva, es ofrecer objetos textuales digitales; en sus inicios como bibliotecas locales por medio de soportes como el CD-ROM y posteriormente en línea. Sin embargo, documentalistas y biblioteconomistas establecen diferencias entre unas y otras. Así Hípola, Vargas-Quesada y Senso (2000: 6) aclaran que la biblioteca digital “implica el uso de las telecomunicaciones”, mientras que la

biblioteca virtual “suele estar formada por un conjunto de conexiones a varias fuentes de información recopilada en internet”. Recientemente Bazán Bonfil (2015) establece que la biblioteca virtual se centra en la digitalización de testimonios mientras que la “digital no ofrece al usuario un objeto conservado en una institución, sino los textos transmitidos mediante” esos objetos. En cualquier caso, el ejemplo más antiguo de estas bibliotecas es el Project Gutenberg (<http://www.gutenberg.org>), creado en 1971 por Michael S. Hart y cuyos fondos lo constituyen básicamente obras en el dominio público, es decir, textos cuyos derechos de autor habían caducado en los Estados Unidos y los ofrecen en un formato de texto plano para que así sean adaptables a cualquier máquina y sistema que existiera o pueda existir en el futuro.

En este contexto, surge en 1999, auspiciada por la Universidad de Alicante y el Banco Santander, la Biblioteca Virtual Miguel de Cervantes (BVMC en adelante: <http://www.cervantesvirtual.com>), que se ha convertido en la gran biblioteca virtual del español. Tal y como informan Bia y Pedreño (2001), la BVMC cubre muchas áreas, desde una fonoteca a tesis doctorales, aunque precisan que “the vast majority of our present digital books are public-domain Hispanic classics, from the twelfth-century to present” (164). La incorporación de los materiales se hace por medio de uno de estos dos métodos: 1) portales especializados y 2) libros tradicionales.

Es este segundo método, y grupo de materiales, el que interesa en este estudio. Lo definen como “the Hispanic classics mentioned above [en la pág, 164] with almost no multimedia material (except for some books that may contain static images)” (Bia y Pedreño, 2001: 168). Es decir, la BVMC básicamente contiene obras *clásicas* hispánicas que se encuentran en dominio público. Eso, en un principio, implica que no pueden incorporar ningún texto editado ni publicado que se encuentre sujeto a derechos de autor. Esto ha provocado las más duras críticas contra la BVMC puesto que el valor *científico* de los textos incorporados es altamente dudoso ya que no son filológicamente fiables (Lucía Megías, 2009: 19; 2012: 105–106), y parece ser una constante de las bibliotecas digitales observada por otros investigadores (Apollon, Bésilile y Régner, 2014; Shillingsburg, 1999: 161).

Bia y Pedreño (2001: 169, fig. 5) informan con cierto detalle del proceso por medio del cual un texto, una obra dada, un libro tradicional, por usar su propia expresión, entra a formar parte de los fondos de la BVMC. Se reduce cuatro pasos básicos:

1. Escaneo (digitalización) y reconocimiento óptico de caracteres de la obra.
2. Corrección y etiquetado TEI-XML.
3. Catalogación.
4. Publicación en la web por medio de la transformación de los ficheros XML a través de hojas de estilo (XSLT) que llegan a producir páginas web (en HTML) o ficheros en formato pdf e incluso en ePub, como se verá.

De una obra clásica cualquiera en la actualidad ofrecen:

1. Ficha (fig. 1).
2. Un acceso a la versión digital de la obra "Leer la obra" (puede ser en formato HTML y por tanto reutilizable, [-http://bit.ly/1AH6W7d-](http://bit.ly/1AH6W7d) o una reproducción facsimilar página a página [-http://bit.ly/1A3Qpq1-](http://bit.ly/1A3Qpq1) y, por tanto, no reutilizable, aunque el icono de acceso indique HTML).
3. Concordancia (no existe en el caso de ediciones facsimilares; fig. 2) [-http://bit.ly/1DP3aMR-](http://bit.ly/1DP3aMR).

Todo esto, a pesar de las críticas de no fiabilidad científico filológica de los textos y de que se haya primado el criterio de cantidad sobre el de calidad, hace que la BVMC sea una poderosa fuente de información que ha ido, en algunos aspectos, evolucionando a bien en los dieciséis años de vida.

The screenshot shows the BVMC website interface. At the top, there is a search bar with the text "libro de los gatos" and a "Buscar" button. To the right of the search bar are social media icons for Twitter, Facebook, YouTube, and RSS, along with a link to "el Blog". The main header of the website reads "BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES" with the website URL "www.cervantesvirtual.com". Below the search bar, there is a "Búsqueda avanzada >>" link.

The main content area features the article title "La alabanza, o que me prohíban éste / Mariano José de Larra". To the left of the article title is a small image of the book cover. Below the title, there is a list of metadata:

- Autor:** Larra, Mariano José de, 1809-1837
- Título:** La alabanza, o que me prohíban éste / Mariano José de Larra
- Publicación:** Alicante : Biblioteca Virtual Miguel de Cervantes, 2002
- Portal:** Mariano José de Larra
- Encabezamiento de materia:** [Literatura periodística española -- Siglo 19º](#)
- CDU:** 821.134.2-92™18"
- Idioma:** español
- URI:** <http://www.cervantesvirtual.com/nd/ark:/59851/bmcm278>

 Below the metadata, there are two icons: "HTML" and "Concordancias".

To the right of the article title, there is a section titled "MÁS RECIENTES" which lists several other articles, including:

- Algunas reflexiones metodológicas sobre la Historia Constitucional / Joaquín Varela Suanzes-Carpegna
- La trayectoria intelectual y política de Adolfo Posada / Joaquín Varela Suanzes-Carpegna
- Diego Ignacio de Córdoba y el papel de Madrid en el mercado crediticio en la Castilla del siglo XVII / Máximo Diago Hernando
- La reforma del Monestir de Monserrat a finales del segle XVI. Epistolari del Monjo i Abat Fra Andrés de San Román (1562-1589) / Bernat Marçal i Andreu
- El mercader Alonso de Ariánzon lanero y comisionista de seguros marítimos / Manuel Basas
- Miguel de Zamora: la vocación religiosa de un viejo mercader / Manuel Basas
- Los sobrinos de Simón Ruiz / Manuel Basas
- Testamento y bienes del mercader burgalés Vitores Ruiz Embito, hermano de Simón / Manuel Basas
- Testamento y mayorazgos del mercader Simón Ruiz Embito / Manuel Basas
- Tráfico atlántico asegurado en Burgos a mediados del XVI / Manuel Basas

At the bottom of the page, there is a footer with the text "Fundación Biblioteca Virtual Miguel de Cervantes" and "Mapa del sitio | Política de cookies | Marco legal".

Figura 1. Ficha de un artículo de Mariano José de Larra en la BVMC – [http://bit.ly/1Jlab5N-](http://bit.ly/1Jlab5N)

- 4 Debido a la extrema longitud de las url (universal resource locator) de la BVMC y lo complicado que es manejarlas en las manchas tipográficas de las publicaciones, las acorto con la utilidad Bitly™ (<https://bitly.com/> acceso 24.02.15). Todos los enlaces funcionan correctamente en la fecha de remisión a la revista y se han comprobado durante la fase de corrección de pruebas, pero no hay garantías de que continúen funcionando en el futuro.

Ahora bien, no siempre la BVMC utiliza textos y ediciones en el dominio público, pues en un principio los textos medievales y del siglo de Oro se digitalizaron “a partir de ediciones en el mercado, sin indicación de su procedencia para así no tener que pagar derechos a los editores o las editoriales” y debido a las protestas de la comunidad científica “se optó por elegir ediciones ‘sin problemas’” (Lucía Megías, 2009: 19).

The screenshot shows the BVMC website interface. At the top, there is a navigation bar with social media icons and the text 'El Bibliotecario | Ayuda'. Below this is a search bar with the text 'Búsqueda por título, autor o contenido' and a 'Buscar' button. A secondary search option 'Búsqueda avanzada >>' is also visible. The main content area features a red header for the entry: 'Poema de Alfonso Onceno, rey de Castilla y de León'. To the right, there is a 'MÁS RECIENTES' section listing various historical and literary references. The entry details include the title, publication information (Alicante: Biblioteca Virtual Miguel de Cervantes, 2008), a reproduction note, and a list of secondary entries by Janer, Pidal, and Sánchez. The CDU classification is 821.134.2-13*13* and the language is Spanish. The URI is provided at the bottom of the entry.

Figura 2. Ficha del Poema Alfonso Onceno en la BVMC – <http://bit.ly/1AH7Q3y>–

Así, aunque las fichas indiquen en las “notas de reproducción original” que han utilizado una de tales ediciones “sin problemas”, sin embargo, reproducen otra que no está en el dominio público y que, por tanto, está sujeta a derechos de autor. De este modo, se explica que, en ocasiones, las indicaciones del original que han digitalizado sean confusas. Por ejemplo, en la ficha de la *Lozana andaluza* de Francisco Delicado (<http://bit.ly/1FUj8Ga>) se dice que es una “edición digital basada en la edición facsímil de Antonio Pérez Gómez, Valencia: Tipografía Moderna, 1950. Edición original: *Retrato de La lozana andaluza*, Venecia, 1528”, pero en el comienzo del texto puede leerse la siguiente nota preliminar:

presentamos una edición modernizada de La lozana andaluza de Francisco Delicado, Venecia, 1528, (edición facsímil de Antonio Pérez Gómez, Valencia, Tipografía Moderna 1950), basándonos en la edición de Bruno M. Damiani (Delicado, Francisco La lozana andaluza, Madrid, Castalia, 1984), cuya consulta recomendamos. Con el objetivo de facilitar la lectura del texto al público no especializado se opta por ofrecer una edición modernizada y eliminar las marcas de editor, asumiendo, cuando

lo creemos oportuno, las correcciones, reconstrucciones y enmiendas propuestas por Damiani (<http://bit.ly/1w6Nbh>, acceso 24.2.15).

Luego ¿cuál es la edición que han digitalizado: la de Venecia de 1528, el facsímil de la misma de 1950 o la de Damiani de 1984?

Desde 1999, según declara la ficha (fig. 3), la BVMC tiene colgada una edición el *Libro de la caza de las aves* de Pero López de Ayala y aclaran en el apartado “notas de reproducción original” que se trata de una “Edición digital a partir de la de José Gutiérrez de la Vega, Madrid, M. Tello, 1879, (Biblioteca Venatoria; 3)”. En el cuerpo del texto, en la primera “página”, explicitan en la nota preliminar que es una:

edición digital a partir de la de José Gutiérrez de la Vega (Madrid, M. Tello, 1879. Biblioteca Venatoria, T. III) y cotejada con la versión de José Fradejas Lebrero (Madrid, Castalia, Col. Odres Nuevos, 1969), cuyos criterios seguimos en la mayoría de los casos. Recomendamos el estudio preliminar de dicha edición para la correcta valoración crítica de la obra (–<http://bit.ly/1ApXAxU>, acceso 24.2.15).

Cabe plantearse ¿qué edición del *Libro de la caza de las aves* ha utilizado la BVMC: la de José Gutiérrez de la Vega (JGV en adelante) de 1876 y, por tanto en dominio público, o la de José Fradejas Lebrero (JFL en adelante) de 1969 y, por tanto, protegida por derechos de autor?

The screenshot shows the BVMC website interface. At the top, there is a search bar and navigation links. The main content area displays the record for 'Libro de la caza de las aves' by Pedro López de Ayala. The record includes the following information:

- Autor:** López de Ayala, Pedro, 1332-1407
- Título:** Libro de la caza de las aves / Pedro López de Ayala
- Publicación:** Alicante : Biblioteca Virtual Miguel de Cervantes, 1999
- Publicación original:** Madrid, M. Tello, 1879
- Notas de reproducción original:** Edición digital a partir de la de José Gutiérrez de la Vega, Madrid, M. Tello, 1879, (Biblioteca Venatoria; 3).
- Portal:** BIBLIOTECA UNIVERSITARIA. Servicio de Información Bibliográfica y Documental de la Universidad de Alicante. SiBID
- Encabezamiento de materia:** Caza -- Siglo 14^o
- CDU:** 639.1"13" 821.134.2-9"13"
- Idioma:** español
- URI:** <http://www.cervantesvirtual.com/nd/ark:/59851/bmch41n3>

Below the record, there are links for 'Leer obra' and 'Concordancias'. To the right, there is a section titled 'MÁS RECIENTES' with a list of recent works, including 'Algunas reflexiones metodológicas sobre la Historia Constitucional / Joaquín Varela Suanzes-Carpegna', 'La trayectoria intelectual y política de Adolfo Posada / Joaquín Varela Suanzes-Carpegna', 'Diego Ignacio de Córdoba y el papel de Madrid en el mercado crediticio en la Castilla del siglo XVII / Máximo Diago Hermandó', 'La reforma del Monasterio de Monserrat a finales del siglo XVI. Epistolari del Monje / Abat Fra Andrés de San Román (1562-1589) / Bernat...', 'El mercader Alonso de Aranzón lanero y comisionista de seguros marítimos / Manuel Basas', 'Miguel de Zamora: la vocación religiosa de un viejo mercader / Manuel Basas', 'Los obreros de Simón Ruiz / Manuel Basas', 'Testamento y bienes del mercader burgalés Vitoroz Ruiz Embito, hermano de Simón / Manuel Basas', 'Testamento y mayorazgos del mercader Simón Ruiz Embito / Manuel Basas', and 'Tráfico atlántico asegurado en Burgos a mediados del XVI / Manuel Basas'.

Fig. 3. Ficha del *Libro de la caza de las aves* en la BVMC – <http://bit.ly/1DRu45e>, acceso 24.2.15–

Una lectura en diagonal y comparada del texto colgado por la BVMC y del de la edición de JGV muestra que el texto digitalizado no es lo indicado. Así, por ejemplo, el texto de la BVMC subdivide los capítulos con una serie de subepígrafes –Causas, síntomas, recetas (cap. X; <http://bit.ly/1Doh2Oc>); síntomas, remedio, complicaciones, síntomas, receta, píldoras, medicación, otra complicación, curación (cap. XI; <http://bit.ly/1AHMtfI>); causas, síntomas, recetas, recomendaciones (cap. XII; <http://bit.ly/1GSy7km>); causas y síntomas, recetas varias, receta de reposo (cap. XIII; <http://bit.ly/1BGWs8j>)– que están totalmente ausentes en el texto de JGV (Gutiérrez de la Vega, 1879: 217–231). Véase como aparece el capítulo X en JGV y en BVMC:

Gutiérrez de la Vega (1879: 216–218)	BVMC (http://bit.ly/1Doh2Oc)
CAPÍTULO X. <i>Cómo se debe purgar el falcon del agua común que no es vidriada.</i> ACAESCE muchas veces, segund dicho hemos, que los mercadores compran et ayuntan falcones para vender, et non curan de al sinon de los gobernar á la menor costa que pueden. Et por tanto non les dan si non malas viandas, et por esto, et por estar encerrados que non veen el sol nin les prueban agua, non están sanos, et cárganse de agua. Otrósí, bien así cuando los cazadores los traen por les non dar á tirar, ó les dar casa con fumo ó sereno, cárganse de agua, et esta agua es ligera de curar antes que se vedrie, et conoscerlo hás en esto: que le verás cuando le dás de comer et tira, que le cae agua por las narices, et estornuda, en guisa que al cazador que le dá de comer rocía el rostro con el agua que sacude. Et si el agua que tú vieres es tan cargada que tiene las señales que dice en el capítulo onceno, que es el siguiente deste, que fabla del agua vidriada, farás et curarlo hás como allí dice.	Capítulo X Cómo se debe purgar el halcón del agua común que no es vidriada <i>Causas:</i> Acaece muchas veces, según hemos dicho, que los mercaderes compran y juntan halcones para vender y no cuidan de otra cosa, sino de alimentarlos a la menor costa que pueden. Por tanto, no les dan sino malas viandas, y por esto, y por estar encerrados, que no ven el sol ni les prueban el agua, no están sanos y cárganse de agua. También, cuando los cazadores los traen y no les dan a tirar, o les dan casa con humo o humedad, cárganse de agua, y es ligera de curar antes de que se vidrie, y lo conocerás en esto: <i>Síntomas:</i> Que le verás cuando le das de comer y tira, que le cae agua por las narices y estornuda, en guisa que al cazador que le da de comer rocía el rostro con el agua que sacude. Y si el agua que tu vieres es tan cargada que tiene las señales que dice el capítulo onceno, que es el siguiente de éste, que habla del agua vidriada, habrás de curarlo como allí dice.

<p>Gutiérrez de la Vega (1879: 216–218)</p>	<p>BVMC (http://bit.ly/1Doh2Oc)</p>
<p>Pero si non es tanta como dicho hé en este capítulo, dale favarraz bien mondado et limpio, apretado en un paño en el agua caliente, en guisa, que salga tan claro que apenas tenga leche; et ponle en cada ventana cuatro gotas, ó cada tres, segund vieres la complision del falcon, et muéstrale un poco el sol, et tíralo luego á la sombra, et esté quedo en una alcándara fasta que faga sus babadas, et ponlo de sí en una cámara fria, et bien tarde dále de comer de una pierna de polla, et para bien mientes cuando así hobieres á purgar tu falcon, que esté bien recio, ca de otra manera sería grand peligro. Et otros dejan de darle el favarraz, et úntanle al falcon los paladares con la miel, et después frégangelos con oruga molida, et fácenle purgar del agua, et es mas sin peligro, et por escusar dende adelante siempre desto, dá siempre á tu falcon á tirar et desplumar dos veces al día, et guárdalo de le non dar sereno de noche ó fumo ó mala vianda, et así nunca se cargará de agua para que lo hayas de melecinar.</p>	<p><i>Recetas:</i> Pero si no es tanta como he dicho en este capítulo, dale favarraz bien mondado y limpio, apretado en un paño en el agua caliente, en guisa que salga tan claro que apenas tenga leche; y ponle en cada ventana cuatro gotas, o, en cada, tres, según vieres la complexión del halcón, y muéstrale un poco el sol y ponlo luego a la sombra, y esté quedo en una alcándara hasta que haga sus babadas, y ponlo más tarde en una cámara fría, y bien tarde dale de comer de una pierna de polla, y para bien mientes, cuando así hubieres de purgar tu halcón, que esté bien recio, porque de otra manera sería gran peligro. Otros dejan de darle el favarraz, y untan al halcón el paladar con miel, y después se lo refriegan con oruga molida, y hácenle purgar del agua, y es más sin peligro. Para excusar esto en adelante, da siempre a tu halcón a tirar y desplumar dos veces al día, y guárdalo que no le dé sereno de noche, o humo, o mala vianda, y así nunca se cargarán de agua para que lo hayas de medicinar.</p>

Se puede aceptar como parte del trabajo de preparación editorial para la publicación dentro de la BVMC una cierta modernización en la grafía (sustitución de <f>– por <h>– (*falcon* > *halcón*, *faga* > *haga*, *fumo* > *humo*), división en párrafos, cambios de acentuación (*hás* > *has*, *ó* > *o*, *falcon* > *halcón*), modernización léxica (*et* > *y*, *ayuntan* > *juntan*, *otrosí* > *también*, *mercadores* > *mercaderes*, *melecinar* > *medicinar*). Sin embargo, los cambios que, aparentemente, han introducido los “correctores” de la BVMC son tales que vienen a coincidir, según parece, *verbatim* con el texto de la edición de JFL:

JFL (López de Ayala, 1969: 99–100)	BVMC (http://bit.ly/1Doh2Oc)
CAPÍTULO X cómo se debe purgar al halcón del agua común que no es vidriada	Capítulo X Cómo se debe purgar el halcón del agua común que no es vidriada
<i>Causas:</i> Acaece muchas veces, según hemos dicho, que los mercaderes compran y juntan halcones para vender y no cuidan de otra cosa, sino de alimentarlos a la menor costa que pueden. Por tanto, no les dan sino malas viandas, y por esto, y por estar encerrados, que no ven el sol ni les prueban el agua, no están sanos y cárganse de agua. También, cuando los cazadores los traen y no les dan a tirar, o les dan casa con humo o humedad, cárganse de agua, y es ligera de curar antes de que se vidrie, y lo conocerás en esto:	<i>Causas:</i> Acaece muchas veces, según hemos dicho, que los mercaderes compran y juntan halcones para vender y no cuidan de otra cosa, sino de alimentarlos a la menor costa que pueden. Por tanto, no les dan sino malas viandas, y por esto, y por estar encerrados, que no ven el sol ni les prueban el agua, no están sanos y cárganse de agua. También, cuando los cazadores los traen y no les dan a tirar, o les dan casa con humo o humedad, cárganse de agua, y es ligera de curar antes de que se vidrie, y lo conocerás en esto:
<i>Síntomas:</i> Que le verás cuando le das de comer y tira, que le cae agua por las narices y estornuda, en guisa que al cazador que le da de comer rocía el rostro con el agua que sacude. Y si el agua que tu vieres es tan cargada que tiene las señales que dice el capítulo oncenno, que es el siguiente de éste, que habla del agua vidriada, habrás de curarlo como allí dice.	<i>Síntomas:</i> Que le verás cuando le das de comer y tira, que le cae agua por las narices y estornuda, en guisa que al cazador que le da de comer rocía el rostro con el agua que sacude. Y si el agua que tu vieres es tan cargada que tiene las señales que dice el capítulo oncenno, que es el siguiente de éste, que habla del agua vidriada, habrás de curarlo como allí dice.
<i>Recetas:</i> Pero si no es tanta como he dicho en este capítulo, dale favarraz bien mondado y limpio, apretado en un paño en el agua caliente, en guisa que salga tan claro que apenas tenga leche; y ponle en cada ventana cuatro gotas, o, en cada, tres, según vieres la complexión del halcón, y muéstrale un poco el sol y ponlo luego a la sombra, y esté quedo en una alcándara hasta que haga sus babadas, y ponlo más tarde en una cámara fría, y bien tarde dale de comer de una pierna de polla, y para bien mientes, cuando así hubieres de purgar tu halcón, que esté bien recio, porque de otra manera sería gran peligro.	<i>Recetas:</i> Pero si no es tanta como he dicho en este capítulo, dale favarraz bien mondado y limpio, apretado en un paño en el agua caliente, en guisa que salga tan claro que apenas tenga leche; y ponle en cada ventana cuatro gotas, o, en cada, tres, según vieres la complexión del halcón, y muéstrale un poco el sol y ponlo luego a la sombra, y esté quedo en una alcándara hasta que haga sus babadas, y ponlo más tarde en una cámara fría, y bien tarde dale de comer de una pierna de polla, y para bien mientes, cuando así hubieres de purgar tu halcón, que esté bien recio, porque de otra manera sería gran peligro.

JFL (López de Ayala, 1969: 99–100)	BVMC (http://bit.ly/1Doh2Oc)
Otros dejan de darle el favarraz, y untan al halcón el paladar con miel, y después se lo refriegan con oruga molida, y hácenle purgar del agua, y es más sin peligro. Para excusar esto en adelante, da siempre a tu halcón a tirar y desplumar dos veces al día, y guárdalo que no le dé sereno de noche, o humo, o mala vianda, y así nunca se cargarán de agua para que lo hayas de medicinar.	Otros dejan de darle el favarraz, y untan al halcón el paladar con miel, y después se lo refriegan con oruga molida, y hácenle purgar del agua, y es más sin peligro. Para excusar esto en adelante, da siempre a tu halcón a tirar y desplumar dos veces al día, y guárdalo que no le dé sereno de noche, o humo, o mala vianda, y así nunca se cargarán de agua para que lo hayas de medicinar.

Es evidente que en estos pequeños párrafos no hay diferencia alguna entre el texto ofrecido por la BVMC y el de JFL. El primero (BVMC) es copia literal del segundo (JFL).

Puesto en conocimiento de la editorial de la edición de Fradejas Lebrero este hecho, que la BVMC había colgado una copia *verbatim*, la respuesta fue:

Lo que me dice es muestra pero no definitivo; habría que hacer un estudio de todo el texto, en su totalidad, comparando el que está colgado con el nuestro... He consultado con mi departamento legal, y me confirman que hay que anotar todas las coincidencias y evaluar... antes de hablar y preguntar a la BVMC (2011.12.21, comunicación personal).

E indicaron que:

[...] Debería hacerlo un estudioso o profesor de la época/textos cuya palabra de especialista pudiera luego ser válida ante un juez, de modo que buscaríamos a uno que nos pudiera ayudar (2011.12.21, comunicación personal).

Un problema añadido es que la BVMC no solo haya “colgado” en su portal esta copia literal del texto, sino que además, en un acuerdo con la casa Grammata, se ofreció en formato ePub, como complemento gratuito, en la edición 2 del *e-reader Papyre 6.1* (Fig. 4). Posteriormente, se ha podido constatar que la editorial barcelonesa Linkgua ofrece una edición del *Libro de la caza de las aves* (<http://bit.ly/1FXntbD>, acceso 25.2.15), realizada “a partir de la de José Gutiérrez de la Vega, Madrid, M. Tello, 1879”, que se vende, entre otras tiendas en línea, en Amazon, tanto en formato Kindle (<http://amzn.to/1AJP7UZ>, acceso 25.2.15) como en papel, o tapa blanda, como la designa esta librería en línea (<http://amzn.to/1AretrZ>, acceso 22.5.15), como a través de Google Books (<http://bit.ly/1BuCBfy>, acceso 25.2.15).

3. MÉTODO

El objetivo es demostrar fehacientemente que el texto colgado por la BVMC y el ofrecido por Linkgua (LK en adelante), tanto en formato digital como con “tapa blanda” no son, como alegan, ediciones basadas en la que José Gutiérrez de la Vega publicó en 1879 y que, por tanto, se encuentra en dominio público, sino que todas ellas son copia literal de la edición que José Fradejas Lebrero publicó en 1969⁵, que no está en el dominio público.



Fig. 4. Detalle del embalaje de Papyre 6.1 edition 2 de Grammata.

3.1 PROCEDIMIENTO

Para realizar un análisis tanto cualitativo como cuantitativo de las semejanzas y disimilitudes que pueden existir entre los textos ofrecidos por la BVMC y LK y su supuesta fuente, la edición de JGV, y demostrar que lo han utilizado tanto por la BVMC y como por LK ha sido la versión de JFL se ha seguido el siguiente procedimiento:

1. Obtención una copia del texto de la web de la BVMC.
2. Extracción del texto del ePub que el *e-reader Papyre 6.1 edition 2* incluye.
3. Adquisición de la versión Kindle de Amazon (LK) y extracción del texto.
4. Adquisición de un ejemplar de tapa blanda de LK y extracción del texto.
5. Digitalización de la edición de 1969.
6. Digitalización de la edición de 1879.

Los puntos 1 y 2 fueron rápidos y sencillos de ejecutar. En el primer caso tan fácil como marcar, cortar y pegar en un editor en texto plano, es decir, en uno que no introduzca

⁵ En realidad, lo que siguen todas estas copias es la segunda edición, pues la primera es de 1959.

información relativa al formato (tipo de letra, tamaño, colores, diseño de página, etc.). Para la conversión del ePub hay varias posibilidades, unas más complicadas que otras. La más sencilla es por medio del programa de gestión de bibliotecas digitales personales Caliber 2.20⁶ que convierte el fichero de partida en otro de formato texto plano en pocos segundos. La salida fue un fichero .txt que presentó un pequeño problema: exceso marcas de tabulación, pero fáciles de retirar con cualquier procesador de texto.

El paso 3 fue un poco más complicado. La arquitectura de Kindle no permite manejar libremente el fichero (¡para evitar la piratería!) ni convertir a otros formatos como en el caso del ePub de Grammata ya que tienen protección DRM (Digital Rights Management), pero se puede leer en otros aparatos que no sean Kindle, lo que permite otras posibilidades, aunque más laboriosas. En este caso se utilizó un iPad Mini 2 con el que se realizaron pantallazos de cada página y se guardaron como imágenes .png (Portable Network Graphics). Posteriormente se procesaron con Drop2Read, un programa de reconocimiento de texto incluido en los escáneres IRIScan, y se obtuvo un fichero .rtf (Rich Text File).

Los pasos 4, 5 y 6 son los más largos y complicados. En todos estos casos se digitalizaron los ejemplares con un escáner Canon iR-ADV C2030i, se procesaron con Drop2Read y se guardaron como ficheros .rtf. En este caso los problemas de error del OCR son algo elevados y ha requerido en todos los casos una lectura y corrección manual de los ficheros.

3.2 ANÁLISIS CUANTITATIVO

Una vez que se convirtieron todos los ficheros a texto plano con codificación UTF-8. Se hizo un primer análisis con Concordance⁷ y AntCon⁸ con el que se obtuvo una primera visión cuantitativa de los textos (Tabla 1).

	JGV	JFL	BVMC	LK
palabras tipo	4306	4332	4334	4330
palabras token	37952	36267	36267	35784
ratio token/tipo	8,8137	8,3719	8,3680	8,2642
caracteres	161573	155825	155827	153613

Tabla 1. Distribución de la ediciones del *Libro de la caza de las aves*

JFL y BVMC son dos textos numéricamente idénticos ya que la diferencia entre ambos es de 2 palabras tipo y 2 caracteres más a favor de la BVMC. En el caso del texto de LK

6 <http://calibre-ebook.com>, acceso 25.2.15.

7 <http://www.concordancesoftware.co.uk>.

8 <http://www.laurenceanthony.net/software.html>.

nos encontramos con un texto ligeramente más breve: 4 palabras tipo menos que su supuesto original, BVMC y 483 palabras token menos que el par JFL/BVMC. Un análisis visual de los textos impresos (JFL y LK y la versión en línea de BVMC) permitió descubrir que esta diferencia se debe a que el texto de LK convirtió la tabla de capítulos que existe entre el prólogo de la obra y el primer capítulo en el índice del libro. Recuperado este material y reubicado en el lugar adecuado se aproximan bastante las cifras de la serie JFL/BVMC/LK (Tabla 2), aunque persiste una ligera diferencia.

	JGV	JFL	BVMC	LK
palabras tipo	4306	4332	4334	4338
palabras token	37952	36267	36267	36248
ratio token/tipo	8,8137	8,3719	8,3680	8,3559
caracteres	161573	155825	155827	155747

Tabla 2. Distribución una vez corregida la omisión de LK

A pesar de haber añadido la tabla que LK eliminó, su texto es 19 palabras más breve que los de JFL y BVMC. Esto es sencillo de explicar: el índice que la edición de LK ofrece se generó automáticamente a partir de las rúbricas de cada capítulo y estas son ligeramente diferentes de las que se ofrecen del “índice de capítulos” en JFL (López de Ayala, 1969: 55–57) y BVMC (<http://bit.ly/1EkkXGX>), como puede verse en los siguientes casos expuestos en la tabla 3.

	JFL / BVMC “índice de capítulos”	JFL / BVMC rúbricas	LK índice y rúbrica
VIII	Cómo se debe regir y gobernar el halcón neblí y ciertas reglas de práctica	Cómo se debe regir y alimentar el halcón neblí y ciertas reglas prácticas para ello	Cómo se debe regir y alimentar el halcón neblí y ciertas reglas prácticas para ello
XIII	Del halcón tuberculoso	Del halcón que deseca	Del halcón que deseca
XXXIV	De la herida del halcón, así abierta como cerrada	De la herida del halcón	De la herida del halcón
XXXV	De la caída o abatidura del halcón en que se hiere	De la abatidura del halcón	De la abatidura del halcón

Tabla 3. Rúbricas seleccionadas del “índice de capítulos” y rúbricas de los capítulos

Por lo tanto, se procedió a eliminar el “índice de capítulos” que presentan JFL y BVMC al final del prólogo y se procesaron de nuevo con Concordance. Los resultados pueden observarse en la tabla 4.

	JGV	JFL	BVMC	LK
palabras tipo	4299	4329	4330	4330
palabras token	37474	35789	35790	35784
ratio token/tipo	8,7169	8,2673	8,2656	8,2642
caracteres	159403	153635	153637	153613

Tabla 4. Resultados tras eliminar el "índice de capítulos"

El que los tres textos examinados contengan prácticamente el mismo número de palabras no implica, necesariamente que sean idénticos. El siguiente proceso ha sido establecer el porcentaje de palabras idénticas que comparten unos y otros. Para este proceso se ha utilizado el programa CopyCatch Gold que sirve para determinar el grado de similitud entre dos documentos y se emplea normalmente para la detección del plagio en centros educativos y también en los análisis para determinar la autoría de un texto dubitado o si hay plagio (Turell, 2005: 275–298) y que en la versión CopyCatch utilizó Turell (2005) para analizar un caso de plagio entre traducciones de *Julio César* de William Shakespeare.

En este paso uno de los problemas con el que nos hemos enfrentado ha sido el que el software está diseñado y pensado con el inglés como única lengua de trabajo por lo que el tratamiento de las vocales acentuadas (á, é, í, ó, ú y ü) y la ñ ha sido errático. Puesto que lo que se analiza es el grado de similitud entre una serie de textos, y los programas lo que hacen es comparar la igualdad o desigualdad de las palabras que lo constituyen, y dado que se entiende por palabra cualquier secuencia de caracteres alfanuméricos entre dos espacios en blanco (Barrón-Cedeño, Vila y Rosso 2014: 126), se reemplazaron las letras acentuadas en sus correspondientes sin tildes puesto que no tienen ninguna influencia para el análisis ya que no se trata de averiguar cuántas veces puede aparecer el artículo masculino *el* y cuántas el pronombre personal *él*, o si se trata del sustantivo *cántara*, el pretérito imperfecto de subjuntivo *cantara* o el futuro de indicativo *cantará* puesto que la relevancia de esa información es nula en esta fase. La ñ ha sido sustituida por el dígrafo *-nw-*, secuencia imposible en español⁹.

9 Tiene efectos sobre el número de caracteres que contiene el texto, pero es un dato de nula incidencia en esta fase del análisis.

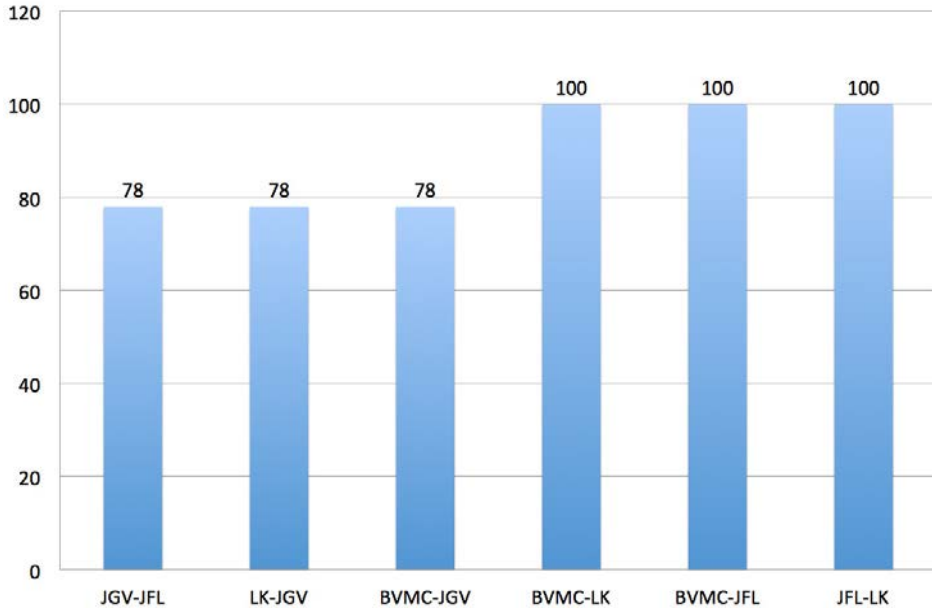


Fig. 5 Porcentaje de vocabulario coincidente en los pares de versiones comparadas

Este análisis ha ofrecido como resultado que entre el texto de JGV y el de la BVMC hay un grado de similitud del 78%, y lo mismo encontramos entre JGV y LK y entre JGV y JFL. Según Turell (2005: 281), es bastante normal que el parecido entre dos textos sobre un mismo tema oscile entre el 30% y el 50% y que “cualquier parecido por encima del 50 %, apuntaría hacia la existencia de algún tipo de relación entre los textos analizados”. Ahora bien, ya que todos los textos que se analizan (JFL, BVMC y LK) alegan ser derivados (actualizados) del texto JGV se podrían considerar y tratar como traducciones de un mismo texto (Fraidejas Rueda, 1991: 49; Jauralde Pou, 1981: 167). En este sentido, Turell (2005: 281) indica que “el nivel umbral del vocabulario coincidente para poder determinar las relaciones textuales existentes será del 70%”, por lo tanto, cuanto más sobrepase este límite más similares serán dos textos, o lo que es lo mismo, mayor será la relación de igualdad que hay entre ellos.

La figura 5 muestra el porcentaje de vocabulario coincidente en más de un caso (coincidente > 1). Los resultados muestran que la tríada JFL / BVMC / LK coinciden con su supuesto original (JGV) en un 78 % lo cual es esperable ya que JGV presenta una versión *medieval* frente a JFL / BVMC / LK que son *traducciones* al español moderno. Lo sorprendente es que estos tres textos derivados coincidan entre sí al 100 %.

La otra medida que se suele utilizar para mostrar el nivel de similitud entre unos textos y otros que son *traducciones* de un mismo original, es decir de JGV, es analizar el número de palabras compartidas una sola vez (compartidas = 1), es decir, todas aquellas palabras

que aparecen una sola vez en cada uno de los textos y que ocurren en los dos miembros de la pareja comparada. Esto es lo que se recoge en la figura 6. Según expone Turell (2005: 282), “cuanto más elevado sea [el] número, tanto [la] puntuación en bruto como la proporción de *vocabulario coincidente*, mayor será el parecido entre los dos textos”. Y eso es lo que sucede en nuestro caso: las puntuaciones más altas se dan entre las parejas de la tríada JFL / BVMC / LK que entre cualquiera de estos tres con respecto a JGV.

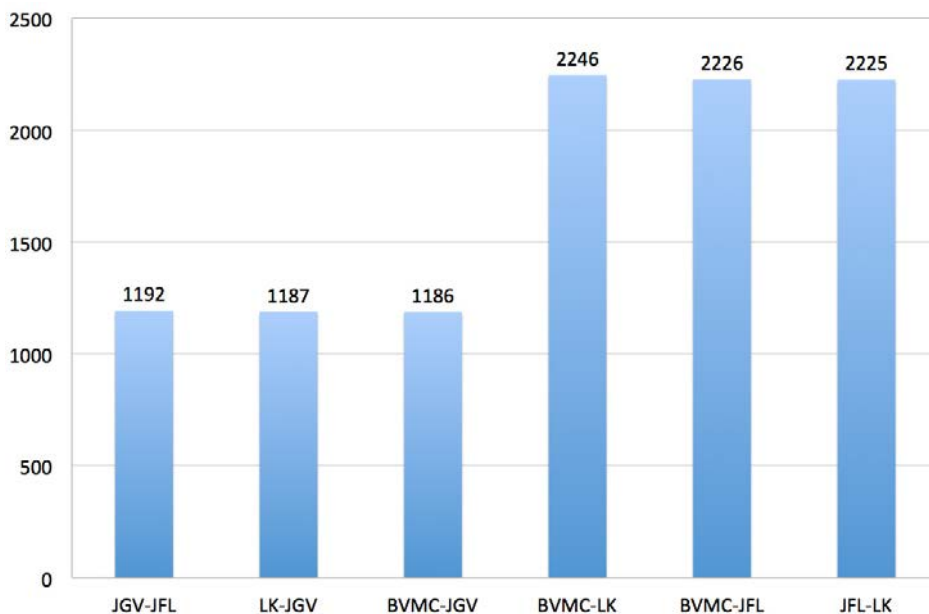


Fig. 6 Palabras compartidas una sola vez

Al ser las versiones de JFL, BVMC y LK, teóricamente, elaboraciones independientes a partir del texto de JGV, sería de esperar un cierto grado de diferencia entre ellas, es decir, que no hubiera habido una semejanza al 100% como ha mostrado el análisis con CopyCathGold (fig. 5). Para probar que debería existir una cierta diferencia entre ellos hemos usado como texto de control la nueva versión modernizada que se ha preparado para sustituir la de JFL de 1969 que identificamos con NON (Fradejas Rueda, en prensa). Esta no se ha creado a partir del texto de JGV, como todas las otras, sino que se trata de una versión realizada *ex novo* que toma como texto base el manuscrito Krahe (Fradejas Rueda, 1989).

	JGV	JFL	BVMC	LK	NON
palabras tipo	4299	4329	4330	4330	4271
palabras token	37474	35789	35790	35784	36620
ratio token/tipo	8,7169	8,2673	8,2656	8,2642	8,5741
caracteres	159403	153635	153637	153613	158649

Tabla 5. Distribución comparativa con el texto de control NON

En primer lugar, NON se ha procesado tanto con Concordance como con AntConc para extraer los mismos datos básicos que en los demás textos examinados (tabla 5) y muestra que hay ciertas diferencias numéricas entre ellos.

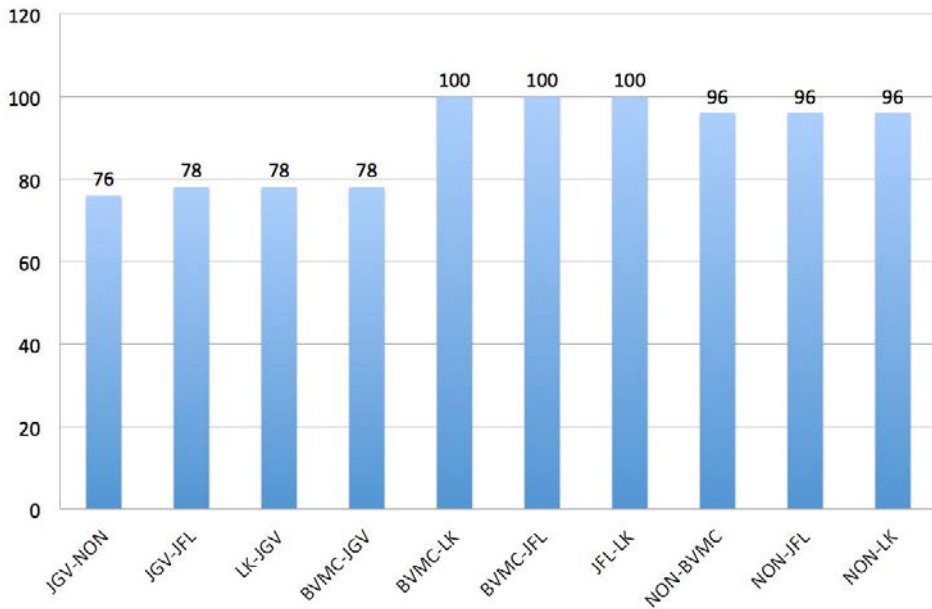


Fig. 7 Porcentaje de vocabulario coincidente en los pares de versiones comparadas añadido el texto de control

En segundo lugar, hemos comparado por medio de CopyCatch Gold la versión NON con la de JVG y el grado de similitud entre ellas es del 76%, lo cual viene a coincidir con el mismo grado de igualdad, aunque ligeramente menor, que hallamos entre JVG y la tríada JFL / BVMC / LK. En tercer lugar se ha comparado NON con la serie JFL / BVMC / LK y el resultado ha sido del 96 % de semejanza, es decir, una ligera disimilitud (figura 7), lo cual implica que si los textos de BVMC y LK fueran elaboraciones independientes a partir de JGV debería de haber, por lo menos, una disimilitud equiparable entre ellos con respecto a JFL como la existe entre ellos y el texto de control NON. En lo que abunda el número de palabras compartidas una sola vez (compartidas = 1) y que se refleja en la figura 8.

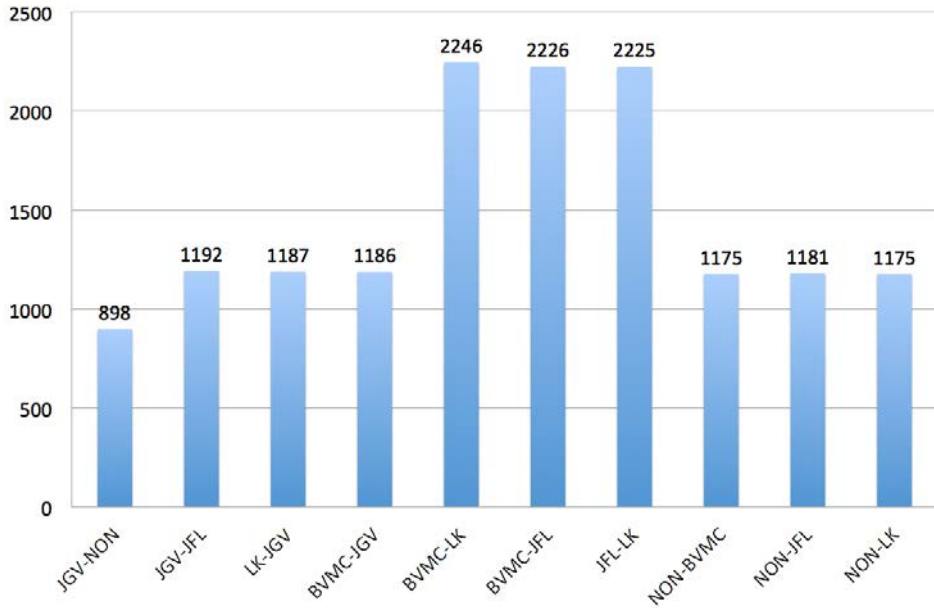


Fig. 8 Palabras compartidas una sola vez añadido texto de control

Sin embargo, el grado de igualdad que ofrece el análisis de JFL, BVMC y LK entre ellos (JFL vs. BVMC; JFL vs. LK; BVMC vs. LK) es del 100 %, lo que viene a coincidir con los datos ofrecidos por Concordance (tabla 5), pues las mínimas diferencias que se dan –una palabra tipo de más en BVMC / LK una palabra token de más en BVMC y cuatro palabras token menos en LK– son despreciables y pueden venir provocadas por el reemplazo de las vocales acentuadas por sus correspondientes sin tilde, por lo que procede averiguar a qué se deben estas mínimas diferencias.

Por otra parte, los programas de concordancias pueden generar listas de palabras, calcular su frecuencia y ordenarlas de varias maneras. Una de ellas es por orden descendente de frecuencias, lo que se puede ver en la tabla 6. Los datos son bastante elocuentes. Si solo tomamos en consideración las columnas JFL, supuesto texto copiado, y BVMC y LK, presuntos copiadore, se ve que la diferencia entre JFL, BVMC y LK es nula. Sin embargo, si observamos los datos particulares de estas 25 palabras pueden apreciarse mínimas discrepancias en las posiciones 1, 12 y 17, aunque no son idénticas. En la posición 1 JFL y BVMC ofrecen el mismo número de ocurrencias para la conjunción y (2311) mientras que LK presenta 3 menos (2308). En la posición 12 BVMC y LK ofrecen el mismo dato: 415 ocurrencias de *halcón*, frente a JFL con una más (416) y en la posición 17 de nuevo BVMC y LK están emparejados frente a JFL, aunque en esta ocasión de manera diferente que en la posición 12: BVMC y LK tienen 1 ocurrencia más (275) que JFL (274) de la conjunción *si*.

	JGV		JFL		BVMC		LK		NON	
	Pal.	Frec.	Pal.	Frec.	Pal.	Frec.	Pal.	Frec.	Pal.	Frec.
1	et	3288	Y	2311	y	2311	y	2308	que	1718
2	que	1507	De	1553	de	1553	de	1553	y	1667
3	de	1486	Que	1524	que	1524	que	1524	de	1543
4	en	1066	En	999	en	999	en	999	el	1053
5	el	961	El	958	el	958	el	958	la	1004
6	la	936	La	858	la	858	la	858	en	987
7	los	613	Los	614	los	614	los	614	los	681
8	las	525	No	520	no	520	no	520	se	572
9	non	498	Las	488	las	488	las	488	las	532
10	á	466	A	459	a	459	a	459	no	519
11	por	438	Se	431	se	431	se	431	a	458
12	falcon	427	Halcón	416	halcón	415	halcón	415	halcón	435
13	es	323	Por	389	por	389	por	389	con	397
14	del	319	Con	322	con	322	con	322	lo	378
15	con	313	Lo	305	lo	305	lo	305	por	373
16	lo	298	Del	279	del	279	del	279	del	303
17	se	295	Si	274	si	275	si	275	si	282
18	son	283	Cuando	252	cuando	252	cuando	252	como	275
19	sí	274	Es	252	es	252	es	252	es	275
20	como	254	Como	239	como	239	como	239	para	259
21	bien	218	Porque	230	porque	230	porque	230	cuando	246
22	le	213	Son	229	son	229	son	229	le	233
23	muy	207	Muy	210	muy	210	muy	210	son	233
24	así	206	Para	208	para	208	para	208	porque	227
25	ca	201	O	205	o	205	o	205	al	213
Total		15615		14525		14525		14522		14863
		41,66%		40,58%		40,58%		40,58%		40,58

Tabla 6. 25 primeras palabras ordenadas por frecuencia

Si BVMC y LK, como indican sus responsables, hubieran sido realizaciones independientes a partir de JGV, hubiera sido esperable una disimilitud parecida a la que ofrece el texto de NON. Las 25 palabras más frecuentes son prácticamente las mismas pero aparecen en distintas posiciones y con distintas ocurrencias salvo en el caso de *halcón* (*falcon* en JGV) que se mantiene en todos los textos en la misma posición: la 12.

3.3 ANÁLISIS CUALITATIVO

Localizar las mínimas discrepancias que se dan entre JFL, BVMC y LK –que consisten, a la vista de los datos de la tabla 5, en una palabra tipo de más en BVMC / LK, una palabra token de más en BVMC y cuatro palabras token menos en LK– sería factible repasando el listado de frecuencias y descubrir en donde no coinciden, como se ha visto en la tabla 6. Sin embargo, este proceder sería a la par que lento, habría que comparar tres listados con 4330 palabras tipo y localizar las discrepancias, proclive al error de lectura además de que no serviría para explicar en qué residen esas divergencias, tan solo que la conjunción copulativa y aparece 3 veces menos en LK, que JFL tiene una ocurrencia más que BVMC y LK y que estos tienen una más de la conjunción *si* que JFL. Aquí entra en juego la vieja filología, la crítica textual, lo cual confirma la aseveración de Blackwell (2012: 5) de que “there is still no single accepted methodology in FL [= Lingüística forense]”, como bien exponen Jiménez Bernal, Reigosa Riveiros y Garayzábal Heinze (2014: 31)

Lo que está fuera de dudas ... es que se trata de una disciplina [la lingüística forense] multidimensional e interdisciplinar en donde tienen cabida numerosos profesionales y expertos afines o ajenos a la propia Lingüística, pero cada uno de ellos con conocimientos que completan y complementan el trabajo de los otros.

3.3.1 Crítica textual

La crítica textual trata de establecer el texto que pudo haber salido de la pluma de un autor, el texto más cercano posible al perdido original puesto que de la práctica totalidad de las obras de los autores de la antigüedad clásica y del medievo se han transmitido por medio de un número indeterminado de copias. Cada vez que se hacía una copia nueva se introducían nuevos errores, o variantes, por no dar una visión negativa de la labor que los copistas hicieron a lo largo de la historia¹⁰. De esta manera, todos los manuscritos que se conservan de un mismo texto difieren en mayor o menor grado unos de otros y, por consiguiente, del perdido original. Así que si se pueden detectar dos variantes comunes, y se puede afirmar que no se remontan al original y, además, que no se han podido originar independientemente (*errores poligenéticos*), los editores suelen establecer que proceden de un antecedente común. Este es el núcleo del llamado método genealógico o lachmanniano y que se basa en el principio del *error común*. Para ponerlo en funcionamiento con un texto del que se tienen varias copias, el procedimiento es tan sencillo como elegir una de ellas como término de comparación, el *texto base* (el

10 Dain (1997: 46; Trovato, 2014: 52) estableció que viene a darse un error por página, con lo que a lo largo de la historia de la transmisión de un texto, las variantes aumentan en proporción geométrica.

problema más serio en esta fase reside en determinar cuál de todos los testimonios debe ser el *elegido*¹¹, confrontarlo con cada uno de los otros testimonios e ir anotando las diferencias. Este procedimiento, que es básico y fundamental, se conoce con el término de *colación* y tradicionalmente se ha hecho a mano apuntando cada una de las variantes, tanto las sustanciales como accidentales¹², con sumo cuidado¹³.

A principios de la década de 1960, Vinton A. Dearing describió un programa creado para un ordenador *mainframe*, un IBM 7090, que le permitía comparar hasta 99 textos y ya aventuraba que “reasonably flexible machines will in time appear on the market at prices that will make them attractive to universities [...] It will them be possible to do textual comparisons in minutes instead of months” (Dearing, 1962: 20). La evolución de los ordenadores, tanto en potencia de procesamiento como en sencillez de manejo y abaratamiento, ha sido enorme desde entonces. El momento de inflexión ocurrió a finales de la década de 1970 y principios de la siguiente, cuando se popularizaron los ordenadores personales. En ese momento muchos filólogos comenzaron a tener acceso a los ordenadores y trabajar con ellos. Susan Hockey (1980: 149–167) ofreció una reseña de las técnicas de computación utilizadas en crítica textual en su primera época, revisión que reelaboró posteriormente y consideró que “of the many collation system that have been written, two stand because out of their sophistication and functionality” (Hockey, 2000: 126): TUSTEP (**TÜ**binger **S**ystem von **TE**xtextverarbeitungs-**P**rogrammen) (Castrillo Benito, 1992) y *Collate* (Robinson, 1989a, 1989b) y que, además, tenían la ventaja de que se habían diseñado para ser ejecutados en los ordenadores personales, TUSTEP en MS-DOS (Windows) y *Collate* en los Apple Macintosh. Ambos programas siguen existiendo y se han actualizado constantemente, sin embargo, no son programas sencillos de manejar puesto que están diseñados para llevar a cabo todas las fases necesarias para una edición crítica de un texto, incluso la preparación final para la imprenta (TUSTEP) o

11 Pérez Priego (2011: 126) indica que “para este cometido [el de colacionar los testimonios] podemos servirnos de una buena y solvente edición moderna de la obra [...], o de una copia provisional de unos de los testimonios transmitidos. Con frecuencia se acude al texto del mejor manuscrito o impreso conservado, o del más completo o más antiguo, o el de la edición más difundida. Pueden hacerse también unas calas en la tradición y decidimos por el texto del testimonio que provisionalmente nos haya parecido el más autorizado”.

12 Este es un concepto que ha venido desde la crítica textual anglosajona: las variantes sustantivas son las que afectan al significado de un autor o a la esencia de su expresión y las accidentales que son las que afectan básicamente a la presentación formal del texto, es decir, ortografía, puntuación o división de palabras (Greg, 1950: 21), lo que viene a coincidir básicamente con lo que en la tradición española se llaman variantes de lengua.

13 Varios manuales de crítica textual como el Pérez Priego (2011: 126–127) o el de West (1973: 66) informan cómo anotar las variantes. Para una exposición de algunos de los sistemas utilizados véase Fradejas Rueda (1991: 57).

en la red (*Collate*)¹⁴. Sin embargo, diversos centros han ofrecido otros programas menos complicados, y que pueden realizar la colación de manera sencilla y rápida a la par que ofrecer otras informaciones y funcionalidades básicas.

Para este proceso hemos utilizado el programa de colación automática Juxta 1.7 (<http://www.juxtasoftware.org>). Este programa se ha empleado en su versión local (*standalone*) y se ha ejecutado en un ordenador iMac (Core 2 Duo, con 4 GB de RAM, sistema operativo OS X Yosemite y Java 1.6)¹⁵. Juxta permite comparar varios testimonios, bien en formato TXT (texto plano) o codificados en XML (especialmente XML/TEI; Fradejas Rueda, 2009), permutando libremente cuál sea el texto base. Una vez colacionados los textos ofrece varias forma de visualización de los resultados, la más interesante, quizá, sea la carta térmica (*heat map*) que muestra todas las variaciones textuales y permite localizar cualquier unidad textual que varíe en cualquiera de los testimonios. También ofrece un histograma que permite ver la densidad de variantes que se pueda dar con respecto al texto base. Otro recurso muy útil de este programa es que puede generar el aparato crítico en HTML.

Para esta parte del análisis, localizar las variantes que diferencian los tres textos –JFL, BVMC y LK–, se parte de los ficheros en los que se han sustituido las letras con tilde por sus correspondientes sin ellas y se ha eliminado el “índice de capítulos” en JFL y BVMC ya que es una fuente de ruido, como se ha mostrado. El resultado ha sido que las diferencias son de mínima entidad y se reducen a 24 variantes que se recogen en la tabla 7.

	JFL	BVMC	LK
1	amigas	ami gas (http://bit.ly/18B6eid)	ami gas (13)
2	halcón	balcón (http://bit.ly/1GPYfJn)	balcón (16; http://bit.ly/1GXaAPq)
3	raleas	ralcas (http://bit.ly/1AS5xYp)	ralcas (20; http://bit.ly/1BLT38i)
4	Romaña	Romana (http://bit.ly/1HWQNgY)	Romana (26; http://bit.ly/1bda9Do)
5	Echanlos	Echados (http://bit.ly/1CoDwNI)	Echados (28; http://bit.ly/1CozLb5)
6	perdería	perderla (http://bit.ly/18B7amB)	perderla (38; http://bit.ly/1EY2YLV)
7	calvo	talvo (http://bit.ly/1MG5gv)	talvo (39; http://bit.ly/1xawX0P)
8	Así	as! (http://bit.ly/1EsrlsU)	as! (45; http://bit.ly/1xu4e1m)
9	bañe	bafle (http://bit.ly/1Em6HNX)	bafle (45; http://bit.ly/1Ghi8Jj)
10	engrendradas (126)	engendradas (http://bit.ly/1bdbSZm)	engendradas (63)
11	ponla	ponía (http://bit.ly/1GQ0uMx)	ponía (66)
12	uñas	tiñas (http://bit.ly/1BsBpCT)	tiñas (67)

14 Collate dejó de existir hacia el año 2006, después de haber sido complementado con el paquete Anastasia para la publicación de ediciones electrónicas (<http://www.sd-editions.com/about/index.html>). CollateX (<http://collatex.net/about/>) se declara el heredero del viejo programa de Robinson (Robinson, 1989a, 1989b).

15 El programa puede ejecutarse en cualquier máquina que tenga un entorno de ejecución Java.

	JFL	BVMC	LK
13	uñas	tiñas (http://bit.ly/1CoEsBJ)	tiñas (67)
14	Cósele	Cásele (http://bit.ly/1xaAGva)	Cásele (70; http://bit.ly/1O1X7aX)
15	envuelta (136)	envuelta (http://bit.ly/1AS8d8o)	envuelta (70; http://bit.ly/1CqFSgn)
16	viantra (140)	viente (http://bit.ly/18B99Hw)	viente (73)
17	cósesela	cásesela (http://bit.ly/1FuCRen)	cásesela (76; http://bit.ly/1GhjMdZ)
18	cóselo	cáselo (http://bit.ly/1Ghqnf7)	cáselo (70; http://bit.ly/1MI321r)
19	cabrón	cabrán (http://bit.ly/1BLXPCB)	cabrán (82; http://bit.ly/1CqGbYG)
20	año	afío (http://bit.ly/1EYag23)	afío (83)
21	emplumados	empluinados (http://bit.ly/1LmzdrX)	empluinados (84; http://bit.ly/19uL5qS)
22	bornís	boráis (http://bit.ly/1MGD4Nv)	boráis (86; http://bit.ly/1CoBAEQ)
23	así	as! (http://bit.ly/1EsrAJN)	as! (89; http://bit.ly/1xu4e1m)
24	Ovidos	olidos (http://bit.ly/1MGD93L)	olidos (98; http://bit.ly/1DvBjBu)

Tabla 7. Variantes que existen entre JFL y BVMC y LK obtenidas con el programa Juxta. Las cifras entre paréntesis remiten a la página correspondiente en las ediciones impresas; los enlaces dan acceso a las versiones en línea; en el caso de BVMC aparecerá resaltado en rojo; en el de LK marcado en amarillo.

BVMC y LK comparten 24 variantes, aunque no todas ellas son errores, sino que en tres casos, lecturas 10, 15 y 16, mejoran el original ya que son erratas que existían en JFL (*engrendadas* 126; *envuelta* 136; *viantra* 140)¹⁶ y las han enmendado. LK presenta una lectura exclusiva frente a JFL y BVMC que consiste en el cambio de la expresión “de mil trescientos y ochenta y seis” (176) por la cifra 1386 (97; <http://bit.ly/1GX8ZJb>). Todas las demás variantes se explican como variantes producidas por errores del programa de OCR o de corrección de los codificadores del texto de la BVMC (lecturas 10, 15 y 16).

Además de la constatación electrónica de que el texto de la BVMC y LK son una copia *ad pedem litterae* de la edición de JFL, también se pueden ofrecer una serie de detalles, obtenidos por medio de la *vieja* filología, que demuestran que BVMC ni LK hicieron uso de su pretendida fuente, JGV, para preparar sus ediciones. Estos detalles surgieron cuando se hizo el análisis comparativo entre el texto de JFL y los de JGV y Lafuente y Gayangos (López de Ayala, 1869) para determinar si había dependencia o no entre las dos ediciones decimonónicas (Dietrick y Fradejas Rueda, 2012). Uno de ellos es que tanto la edición de Lafuente y Gayangos (López de Ayala, 1869: 16) como la de JGV (1879: 161) leen “et los traen en las *contres* de Alemaña” mientras que la lectura de JFL (López de Ayala, 1969: 64) es “y los traen a las *comarcas* de Alemania”; un poco más adelante JGV (1879: 165) y Lafuente y Gayangos (López de Ayala, 1869: 19) leen “Otros falcones neblís hay que es su plumaje rubio, et la *punta* gruesa ... et la *punta* orlada de amarillo” lo que

16 Estas erratas son innovaciones introducidas en la edición de 1969 puesto que no existen en la de 1959.

en JFL (López de Ayala, 1969: 67) aparece como “otros halcones neblís, cuyo plumaje es rubio y el *pico* grueso ... y el *pico* orlado de amarillo” y tanto BVMC como LK en estos tres lugares leen de acuerdo con lo que dice JFL y no con su pretendido original pues ni BVMC ni LK tienen motivo alguno para trocar *contra* por *comarca* ni *punta* por *pico* por lo tanto no dependen de JGV, si hubieran tenido motivos deberían haber corregido *ope ingenii* todos estos otros casos *deslanado, arábigos, de zahara, feciere, fiesta y concas* (Gutiérrez de la Vega, 1879: 163, 163, 189, 311–312 y 313) que son lecturas absurdas en esos puntos. Las lecturas correctas en todos esos casos son: *cocas, pintas, pintas, deslavados, araniegos, zahareños, ficiere siesta y cocas*¹⁷.

Tan fuerte es la dependencia que, aunque localizaron y corrigieron tres erratas (variantes 10, 15 y 16 de la tabla 7), JFL tiene algunas más que han pasado como lecturas *correctas* a BVMC y, por consiguiente, a LK. Una es un error de anticipación, de manera que en JFL aparece *gracero* donde debería aparecer *garcero* que es lo que presenta JGV (1879: 175) y la otra es el no haber disuelto la fusión de una preposición y un demostrativo *–desta–* (López de Ayala, 1969: 130) y que, por consiguiente, se halla en ambas copias (BVMC y LK).

4. CONCLUSIÓN

Tras haber obtenido unas versiones digitales de todos los textos, haberlos descompuesto y analizado con Drop2Read, Caliber, Concordance, AntConc, Copycatch Gold 2 y Juxta se ha podido demostrar que el texto publicado por la Biblioteca Virtual Miguel de Cervantes (BVMC) no es una digitalización de la edición de José Gutiérrez de la Vega (JGV), como indican, con la que tan solo hay un 76% de similitud, lo cual es esperable ya que se trata de una digitalización de ese texto, sino que es una copia fiel de la versión modernizada publicada por José Fradejas Lebrero (JFL), con la que el grado de la igualdad es del 100%. Además, se ha podido mostrar que la versión publicada por ediciones Linkgua (LK) y vendida a través de Amazon y disponible, parcialmente, a través de Google Books no es una versión creada a partir de la ya mencionada de José Gutiérrez de la Vega (JGV), con la que tan solo tiene el 76% de similitud, sino que esta es una copia literal de la colgada versión por la Biblioteca Virtual Miguel de Cervantes (BVMC), lo cual no solo se prueba por el 100% de igualdad que se da entre el texto de LK y el de la BVMC, sino por una serie de innovaciones que se han originado en el texto de BVMC debido a errores del reconocimiento de caracteres (OCR) que se utilizó para convertir la edición originaria, la de JFL, en otra de digital para la BVMC.

17 Este último aparece en JFL (López de Ayala 1969: 156) como *cuencas* y bajo esta forma en BVMC y Linkgua.

Por lo tanto, ni la edición de la BVMC, ni la de LK, en cualquiera de sus versiones o formatos¹⁸, no son una digitalización de la de JGV, sino que son una copia exacta de la de José Fradejas Lebrero.

REFERENCIAS BIBLIOGRÁFICAS

- APOLLON, D.; BÉSLILE, C. & RÉGNIER, Ph. (2014). *Digital Critical Editions*. Urbana: University of Illinois Press.
- BARRÓN-CEDENO, A.; VILA, M. y ROSSO, P. (2014). "Detección automática de plagio: de la copia exacta a la práfrasis". En *Lingüística forense: la lingüística en el ámbito legal y policial*, Elena Garayzábal Heinze, Miriam Jiménez Bernal y Mercedes Reigosa Riveiros (eds.), 123–152. Madrid: Euphonía Ediciones.
- BAZÁN BONFIL, R. (2015). "Una constante. El sueño de una fonoteca digital". En *Textos, píxeles y bits. Reflexiones sobre la publicación digital*, I. Galina Russell (ed.). México: UNAM (disponible en <http://www.iib.unam.mx/files/Investigacion/Publicaciones/Textos-pixeles-y-bits.epub>).
- BIA, A. y PEDREÑO, A. (2001). "The Miguel de Cervantes Digital Library: the Hispanic Voice on the Web". *Literary and Linguistic Computing* 16/1, 161-177.
- BLACKWELL, S. (2012). "History of Forensic Linguistics". En *The Encyclopedia of Applied Linguistics*. Blackwell Publishing. DOI: 10.1002/9781405198431.wbeal0508.
- CASTRILLO BENITO, N. (1992). *Programas Tustep (TUEbinger System von TEXTverarbeitungs-Programmen): aplicación del tratamiento de textos a la investigación*. Valladolid: Universidad de Valladolid.
- COTTERILL, J. (2012). "Corpus Analysis in Forensic Linguistics". En *The Encyclopedia of Applied Linguistics*. Blackwell Publishing. DOI: 10.1002/9781405198431.wbeal0238.
- COULTHARD, M. & JOHNSON, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. London: Routledge.
- DAIN, A. (1997). *Les manuscrits*. Paris: Diderot.
- DEARING, V. (1962). *Methods of Textual Editing. A Paper Delivered at a Seminar on Bibliography Held at the Clark Library, 12 May, 1962*. Los Angeles: William Andrews Clark Memorial Library University of California.

18 Hay otra versión más en línea, la de Miguel Garci-Gómez en el portal Cybertextos (<http://mgarci.aas.duke.edu/cibertextos/JUAN-MANUEL/LIBRO-CAZA/>) patrocinada por la Duke University. No declara la fuente de la que ha obtenido el texto, pero la comprobación de las variantes recogidas en la tabla 7 demuestran fehacientemente que se trata de una copia –sin tildes– de la de la BVMC. Lo único que hay que realmente achacarle a esta copia es el error grave de atribución: la presentan como el *Libro de la caza* de don Juan Manuel. Es evidente que el editor trató el texto como un mero objeto digital sobre el que construir unas concordancias y ofrecerlas en línea, no tuvo ningún interés filológico, lo cual abunda en la baja calidad de muchas bibliotecas virtuales o digitales.

- DIETRICK, D. A. y J. M. FRADEJAS RUEDA (2012). "Bases para una edición crítica del *Libro de la caza de las aves* de Pero López de Ayala". *Revista de Filología Española* 92/1, 43-79.
- FRADEJAS RUEDA, J. M. (1989). "Los manuscritos del *Libro de la caza de las aves*: intento de un censo y descripción del Ms. Krahe". *Epos: Revista de Filología* 5, 497-504.
- ____ (1991). *Introducción a la edición de textos medievales castellanos*. Madrid: Universidad Nacional de Educación a Distancia.
- ____ (2009). "La codificación XML/TEI de textos medievales". *Memorabilia: Boletín de Literatura Sapiencial* 12, 219-247.
- GREG, W. (1950). "The Rationale of Copy-Text". *Studies in Bibliography* 3, 19-36.
- GUTIÉRREZ DE LA VEGA, J. (1879). *Libros de cetrería de el príncipe y el canciller*. Madrid: M. Tello.
- HÍPOLA, P.; VARGAS-QUESADA, B. y SENSO, J. A. (2000). "Bibliotecas digitales: situación actual y problemas". *El Profesional de la Información* 9/9, 4-14.
- HOCKEY, S. (1980). *A guide to computer applications in the humanities*. London: Duckworth.
- ____ (2000). *Electronic texts in the humanities: principles and practice*. Oxford: OUP.
- HOLMES, D. I. (1998). "The Evolution of Stylometry in Humanities Scholarship". *Literary and Linguistic Computing* 13/3, 111-117.
- JAURALDE POU, P. (1981). *Manual de investigación literaria: guía bibliográfica para el estudio de la literatura española*. Madrid: Gredos.
- JIMÉNEZ BERNAL, M.; REIGOSA RIVEIROS, M. y GARAYZÁBAL HEINZE, E. (2014). "La lingüística forense: licencia para investigar la lengua". En *Lingüística forense: la lingüística en el ámbito legal y policial*, E. Garayzábal Heinze, M. Jiménez Bernal y M. Reigosa Riveiros (eds.), 27-47. Madrid: Euphonía Ediciones.
- LEONARD, R. A. (2005). "Forensic Linguistics: Applying the Scientific Principles of Language Analysis to Issues of the Law". *International Journal of the Humanities* 3/7, 65-70.
- LÓPEZ DE AYALA, P. (1869). *El libro de las aves de caça*. Madrid: SBE.
- ____ (1969). *Libro de la caza de las aves*. José Fradejas Lebrero (ed.). Madrid: Castalia.
- LUCÍA MEGÍAS, J. M. (2009). "La edición crítica hipertextual: hacia la superación del incunable del hipertexto". En *Lecturas y textos en el siglo XXI: Nuevos caminos en la edición textual*, C. Castillo Martínez y J. L. Ramírez Luengo (eds.), 11-74. Lugo: Axac.
- ____ (2012). *Elogio del texto digital: claves para interpretar el nuevo paradigma*. Madrid: Fórcola.
- O'KEEFE, A.; MCCARTHY, M. & CARTER, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- OLSSON, J. (2004). *Forensic Linguistics: An Introduction to Language, Crime, and the Law*. London: Continuum.
- PECORARI, D. (2012). "Plagiarism". En *The Encyclopedia of Applied Linguistics*. Blackwell Publishing. DOI: 10.1002/9781405198431.wbeal0914.

- PÉREZ PRIEGO, M. Á. (2011). *La edición de textos*. Madrid: Síntesis.
- PHILBRICK, F. A. (1949). *Language and the Law: the Semantics of Forensic English*. New York: MacMillan.
- ROBINSON, P. M. W. (1989a). "The Collation and Textual Criticism of Icelandic Manuscripts (I): Collation". *Literary and Linguistic Computing* 4/2, 99–105
- (1989b). "The Collation and Textual Criticism of Icelandic Manuscripts (2): Textual Criticism". *Literary and Linguistic Computing* 4/3, 174–181.
- SHILLINGSBURG, P. (1999). *Scholarly Editing in the Computer Age: Theory and Practice*. Ann Arbor: The University of Michigan Press.
- SVARTVIK, J. (1968). *The Evans Statements: A Case for Forensic Linguistics*. Göteborg: Elanders.
- TROVATO, P. (2014). *Everything you always Wanted to Know about Lachmann's Method: A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Structuralism, Cladistics, and Copy-Text*. Padova: libreriauniversitaria.it.
- TURELL, M. (2005). *Lingüística forense, lengua y derecho : conceptos, métodos y aplicaciones*. Barcelona: Institut Universitari de Lingüística Aplicada Universitat Pompeu Fabra.
- WEST, M. (1973). *Textual Criticism and Editorial Technique Applicable to Greek and Latin texts*. Stuttgart: B.G. Teubner.

NOTA: La revista *Signa*, tanto en este como en otros casos similares, publicará las réplicas bien fundamentadas que este artículo pudiese ocasionar.

Recibido el 22 de octubre de 2015.

Aceptado el 30 de octubre de 2015.