
Editar el Siglo de Oro
en la era digital

—•—

SUSANNA ALLÉS-TORRENT
EUGENIA FOSALBA
(EDS)



Studia Aurea Monográfica

Editar el Siglo de Oro
en la era digital



Studia Aurea Monográfica

Studia Aurea Monográfica, coeditada
por la Universitat Autònoma de Barcelona
y la Universitat de Girona,
es una colección auspiciada por
*Studia Aurea. Revista de Literatura Española y
Teoría Literaria del Renacimiento y Siglo de Oro*

Studia Aurea se fundó, en 2007,
con el propósito de ofrecer un instrumento
de intercambio científico y de colaboración
a los investigadores, y de propiciar
una aproximación supradisciplinar
a los estudios literarios.

Desde 2010, *Studia Aurea Monográfica*,
dirigida por Eugenia Fosalba y María José Vega,
publica volúmenes dedicados a analizar,
desde diversas perspectivas críticas,
los temas y problemas capitales
de la investigación más reciente
sobre las letras altomodernas.

Editar el Siglo de Oro en la era digital



SUSANNA ALLÉS-TORRENT
EUGENIA FOSALBA
(EDS)

Este volumen se publica con la ayuda
del PROYECTO DE INVESTIGACIÓN
Garcilaso en Italia, estancia en Nápoles (2016-2019),
con referencia FFI2015-65093-P del Ministerio
de Economía, Industria y Competitividad.

Comité científico

Sònia Boadas
(Universitat Autònoma de Barcelona)

Claudia Corfiati
(Università degli Studi di Bari)

Rodney Lokaj
(Università Kore di Enna)

María Morrás
(Universitat Pompeu Fabra)

ISBN (pdf) 978-84-128138-3-8



Índice

<i>Prefacio</i> , Eugenia Fosalba	9
<i>Introducción</i> , Susanna Allés-Torrent	11
ANNA BOGNOLO & STEFANO BAZZACO <i>Editar libros de caballerías en la era digital: la Biblioteca Digital del Proyecto Mambrino</i>	19
JESÚS TRONCH PÉREZ <i>Editar en TEI teatro clásico español e inglés en los proyectos ARTELOPE y EMOTHE</i>	49
RAMÓN VALDÉS GÁZQUEZ <i>Por la dignidad del texto. El teatro del Siglo de Oro y de Lope de Vega en la red. Principios ecdóticos y de Humanidades Digitales</i>	71
MERCEDES BLANCO y AUDE PLAGNARD <i>Editar una obra y los testimonios de su recepción. El caso de Góngora</i> ..	109
EUGENIA FOSALBA y SUSANNA ALLÉS-TORRENT <i>La edición crítica digital de la obra poética de Garcilaso de la Vega</i> ...	143
HELENA BERMÚDEZ SABEL, CLARA ISABEL MARTÍNEZ CANTÓN y PABLO RUIZ FABO <i>Procesamiento del lenguaje natural y fijación del texto. Experiencias en torno a la constitución de un corpus diacrónico de sonetos</i>	161
BORJA NAVARRO COLORADO <i>Edición digital enriquecida: un modelo de anotación multinivel para poesía del Siglo de Oro</i>	175

Prefacio

Creo que es de justicia reconocer aquí el enorme esfuerzo que Susanna Allés-Torrent ha vertido en el proyecto Pronapoli, desde la tarde afortunada en que nos reunimos en un elegante bar de copas de Enrique Granados, y encaramadas en nuestros respectivos taburetes, apoyó sobre la mesa su libreta de papel reciclado y con un bolígrafo de tinta azul turquesa empezó a poner orden en mis tan entusiastas como deshilvanadas ideas. Muy buenas cabezas han contribuido a lo que hoy es el proyecto Pronapoli pero puedo asegurar que solo aquel día nuestra plataforma digital, que hoy obtiene un impacto arrollador en las redes, empezó a tomar verdadera forma. Lo mismo está sucediendo con *Soleidad amena*, la edición crítica y digital que estamos llevando a cabo ahora, en la que Susanna Allés-Torrent está de nuevo plenamente implicada: no nos faltaban ideas sobre cómo convertir una erudita edición digital en una herramienta de lectura que acercara con un solo clic la obra de Garcilaso a todos los hogares del mundo, perfectamente integrada en el contexto cosmopolita del que surgió, pero valía la pena compartir nuestras dudas con otros grandes profesionales de las Humanidades Digitales del resto de Europa, para resolverlas de la mejor manera posible y dejar testimonio en estas páginas del interesantísimo debate que surgió, a propósito de estas inquietudes, un soleado día de junio de 2023, en el Aula Francesc Eiximenis de la Universitat de Girona.

Eugenia Fosalba
Barcelona, 31 de enero de 2024

Introducción

Susanna Allés-Torrent

La dimensión digital que hace años tomó tierra en el campo de las Humanidades ha ido transformando todas las disciplinas científicas. Esta transformación se ha percibido de manera dispar, para algunos demasiado lentamente —si se piensa en esos primeros programas de concordancias de décadas atrás—, para otros demasiado de repente sin entender todavía las implicaciones intelectuales de la labor digital. La filología y los estudios literarios han sido, desde los años cuarenta del siglo pasado, disciplinas pioneras en la adopción digital, especialmente desde el punto de vista de las herramientas, los métodos y la práctica, así como la reflexión y su teorización han llegado posteriormente. En el caso de los estudios medievales y del Siglo de Oro peninsulares, hace ya más de dos décadas que vienen surgiendo de manera exponencial proyectos digitales que, si bien nacieron en su momento de necesidades reales y urgentes para llevar a cabo nuevas investigaciones (piénsese en las muchas bases de datos sobre fuentes primarias), han empezado paulatinamente a concebir su metodología científica y sus resultados a partir de lo que las Humanidades Digitales aportan. En el caso de la crítica textual, y más generalmente de la edición de textos, piedra de toque en los estudios hispánicos peninsulares, la transformación digital está todavía en curso. De hecho, pocos casos son tan paradigmáticos como el de la crítica textual, metodología científica para el establecimiento de textos fiables, que debe ahora adaptarse o, quizás mejor dicho, reinventarse a partir de unos principios que habían sido claramente establecidos para su práctica analógica (con escuelas diferentes y manuales canónicos) hacia una práctica digital. Así, en los últimos años, no solo ha habido contribuciones muy significativas sobre edición digital, sino que han ido surgiendo proyectos editoriales que han aceptado el reto de producir nuevos objetos científicos, una nueva manera de concebir y presentar el texto, que se aleja del texto impreso, el modelo que hemos consumido durante seis siglos. De la misma manera que la imprenta supuso una revolución en múltiples niveles (producción, distribución, lectura, acceso, etc.), ahora, cuando hablamos de edición filológica digital (o edición académica digital, siguiendo la traducción inglesa), nos adentramos en realidad en tierra de nadie, como decía

Plutarco, a propósito de la descripción de las geografías desconocidas, “lo de más allá, dunas áridas y plagadas de fieras”.

Este interés por la transformación en la edición de los textos del Siglo de Oro es el que ha guiado la publicación de un nuevo volumen que pudiera convertirse en una pieza más para el engranaje teórico sobre crítica textual y prácticas digitales. Por ello, parecía oportuno ofrecer un escenario en el que se pudieran discutir temas relevantes: los modelos de representación de datos procedentes del análisis filológico y de la compleja fenomenología que conllevan las fuentes primarias con las que trabajamos y su conversión a un formato digital; el modelado de datos y el uso de ciertas metodologías en la codificación textual con ciertos estándares como XML-TEI, o los modelos de presentación de datos y los retos en el diseño de las interfaces. Convenía asimismo abrir la discusión sobre el potencial de los datos textuales codificados para su procesamiento digital y la obtención de visualizaciones, o la adopción de técnicas de lenguaje de procesamiento natural. Incluso, nos interesaba profundizar en cuestiones como los flujos de trabajo editorial, las plataformas de edición, las infraestructuras adoptadas, o los retos de colaboración, la preparación técnica del equipo, la falta de financiación y las consiguientes estrategias adoptadas. Todos estos temas, deseábamos que fueran acompañados de las experiencias reales en diferentes proyectos sobre textos del Siglo de Oro.

Presentamos en este monográfico un conjunto de siete contribuciones que representan una aportación tanto desde el punto de vista de la edición y los estudios del Siglo de Oro como el de las metodologías digitales aplicadas a los textos de dicha época. Más concretamente, podemos establecer tres grupos de experiencias similares, aunque los temas se entrecruzan y solapan irremediamente. Por un lado, las bibliotecas digitales; por el otro, las ediciones de autor concebidas con toda su extensión de archivo digital, y, en fin, técnicas de procesamiento y anotación textual encaminadas al establecimiento y fijación del texto literario.

Así pues, todas las contribuciones giran en torno a materias dispares como los libros de caballerías, el soneto siglodorista, la obra y recepción de Góngora, las comedias Lope de Vega o la obra poética de Garcilaso de la Vega. Debemos señalar que las contribuciones no se centran únicamente en el proyecto en curso, sino que pretenden ser una aportación científica al campo de la crítica textual y a las tecnologías del texto, en general, desde el punto de vista digital y en el marco de las Humanidades Digitales. Todas ellas van más allá de la mera descripción de los proyectos en los que se alberga su investigación. Cada uno de los autores se ha esforzado por ofrecer una contextualización general en el seno de las Humanidades Digitales, en términos de práctica y teoría, y aspiran a su aportación en tanto que los métodos de trabajo sean extensibles a otras experiencias.

En el caso de las bibliotecas digitales, Anna Bognolo (Università di Verona) y Stefano Bazzaco (Università di Verona) presentan los contenidos y objetivos del *Proyecto Mambrino*, así como los desafíos encontrados para la creación de la *Biblioteca Digital* de ediciones digitales. Desde hace ya algunos años, ambos

autores, juntamente con un amplio equipo, investigan los derroteros por los cuales los libros de caballería, siguiendo la senda del *Amadís de Gaula*, viajaron desde España y se recibieron en Italia a lo largo de los siglos XVI y XVII. Un fenómeno que no solo incluye traducciones e impresos de textos españoles en las prensas italianas, sino también obras de nueva factura y continuaciones de originales. Esta serie de textos, que hasta la fecha ha recibido poca atención y no ha sido editada, ha empezado a poblar la *Biblioteca digital del Proyecto Mambrino*. En cuanto al tipo de edición, se adopta como premisa la fidelidad al ejemplar más completo y correcto de cada edición príncipe (descartan, por tanto, el establecimiento de un texto crítico con variantes procedentes de la tradición textual). Proponen, así, un modelo editorial que se encaja en su ya creada plataforma concebida como un archivo digital que contiene variedad de recursos (una base de datos dedicada a un censo de obras, otras destinadas a personajes, lugares y motivos caballerescos). Las ediciones propuestas presentan un diseño tripartito: la reproducción digital de la edición príncipe, el texto editado y presentado de una manera clara y diáfana, y un tercer apartado con los resúmenes de los diferentes capítulos concebidos como apoyo a la lectura. En estos últimos, aparecen los enlaces a los diferentes personajes y lugares que, a su vez, constituyen índices de la obra. La infraestructura web utilizada ha sido *TEI Publisher*, una de las plataformas de codificación y publicación más utilizadas en los últimos años. En la base de su presentación web se halla un trabajo minucioso de codificación en XML-TEI que es la que ha permitido establecer las diferentes secciones y cuestiones de identificación semántica apenas mencionadas. El capítulo, además, aporta discusiones relevantes sobre algunos de los retos de codificación textual para los textos de esta época. También es interesante subrayar que, para el proceso editorial, el equipo ha adoptado métodos de transcripción automatizada, concretamente a través de la plataforma Transkribus, con resultados satisfactorios. Así también, entre sus retos a corto plazo aparecen cuestiones centrales en torno a la interoperabilidad entre repositorios institucionales y de consulta para que las ediciones puedan ser recogidas por los catálogos de las bibliotecas (como el protocolo de distribución OAI-PMH), así como sobre la identificación de los recursos digitales (necesidad de la asignación de ISBN y DOIs).

A continuación, Jesús Tronch Pérez (Universitat de València) pone el énfasis en la idea de los textos como datos, su tratamiento electrónico y codificación en XML-TEI a partir del trabajo realizado en el marco de los proyectos de la Biblioteca Digital de ARTELOPE, que difunde las comedias de Lope de Vega (un total por ahora de 334 títulos), y EMOTHE, que ofrece ediciones modernizadas de obras teatrales europeas (italianas, francesas, españolas, portuguesas e inglesas) compuestas entre 1500 y 1700. Tronch estructura su trabajo a partir de la metodología de la crítica textual digital: captura (de dónde proceden y cómo se han transformado digitalmente los textos), estructuración según las guías directrices de la TEI y, en fin, su publicación en línea. Todo ello sin olvidar, a modo de espina dorsal, el flujo de trabajo compartido con otros colaboradores y que en

ciertos aspectos ha debido adaptarse a los programas adoptados en los albores del proyecto (por ejemplo, el uso de FileMaker). Para llevar a cabo el marcado de los textos, como decimos, tanto ARTELOPE como EMOTHE han utilizado el lenguaje XML siguiendo las directrices de la TEI. En su contribución, se ejemplifica una casuística heterogénea procedente de textos dramáticos y se proponen soluciones de codificación para, por ejemplo, recuperar con fines estadísticos antropónimos, topónimos, referencias mitológicas, oficios, citas literarias, latinismos, elementos escenográficos, vestuario, entre muchos otros. Otro de los aspectos que Tronch trae a colación es el papel del editor digital y la terminología utilizada para referirse a esta nueva figura. Tras barajar diferentes términos a partir del inglés (*encoder*, *digital editor*, *encoding architect*, etc.), repropone el término “editor” pues, aunque su labor sea diferente a la del que establece propiamente el texto crítico, este continúa teniendo un componente intelectual crítico, lejos de acciones mecánicas, que le obligan a hacer juicios interpretativos y tomar decisiones editoriales, sin el cual no le habría sido posible la “codificación crítica” del texto. A este propósito, y más allá de las detecciones de posibles errores en el texto fijado, son especialmente esclarecedores algunos ejemplos de codificación que ponen en evidencia su dimensión hermenéutica, como por ejemplo la labor interpretativa del editor digital a la hora de explicitar y marcar la identificación de los personajes o las didascalías, una tarea que el ordenador no detectararía de forma autónoma. En definitiva, Tronch pone sobre la mesa no solo el papel de esta nueva figura editorial y académica, sino que ahonda en una cuestión crucial en el nuevo debate sobre la epistemología de la crítica textual digital. Una discusión cuyas ramificaciones, en el fondo, tocan de cerca el papel del humanista digital, el crédito y la evaluación científica de la labor digital.

Por su lado, la contribución de Ramón Valdés (Universitat Autònoma de Barcelona) representa un balance crítico, desde una perspectiva diacrónica, sobre la cantidad y la calidad de los textos del teatro del Siglo de Oro en la red. En el centro de su reflexión se encuentra la idea de la “dignidad del texto” que surge de la constatación de la existencia de textos en línea establecidos sin rigor crítico que generan y perpetúan una tradición textual deficiente, utilizada después por los mismos investigadores en sus estudios. Ciertamente, Valdés reconoce la utilidad de la existencia de estos textos en abierto y, para el caso de Lope, aplaude iniciativas como las de ARTELOPE, pues es solo la cantidad la que permite adoptar técnicas digitales para estudios como los de estilística o atribución de autoría, por poner un ejemplo. Pero aún así, insiste, con mucha razón, en el imperativo de producir y publicar textos en digital con el mismo rigor filológico que la misma crítica textual ha establecido desde hace décadas en su formato analógico. Desde el equipo de PROLOPE, reconocen que quizás el caso de Lope de Vega sea uno de los más dramáticos. A lo largo de dos décadas, PROLOPE ha ya publicado unas doscientas cincuenta ediciones críticas en papel (en Milenio y Gredos), y, si bien adoptaron el reto de ponerlas en línea y en abierto hace algunos años, cuestiones de tiempo, economía y la complejidad de

transformar las copias digitalizadas en texto electrónico han impedido hacerlo posible con la rapidez o agilidad deseadas. Además, ya sean las digitalizaciones o incluso los archivos nativamente digitales, la densa formalización de la distribución de la página de estas ediciones dificulta su trasvase digital. Piénsese en toda la información allí contenida: prólogo, estudio, anotación, aparato crítico y variantes lingüísticas, notas onomásticas, listas de erratas, índices, etc. Transformar estas ediciones críticas en ediciones críticas digitales es tarea hartamente ambiciosa por una infinidad de motivos. Aún así, este es el objetivo de la *Biblioteca digital PROLOPE*, que contempla publicar en abierto el corpus completo de las ediciones críticas de las comedias con cada una de sus secciones. De momento, la Biblioteca se lanzará con tres títulos: *El bautismo del Príncipe de Marruecos*, *Servir a señor discreto* y *Los ramilletes de Madrid*. Como Valdés explica, esta tarea resulta altamente laboriosa y, por ello, y como estrategia (quizás) rápida y efectiva, el equipo se halla en vías de exploración ante la posibilidad de crear un repositorio de los textos críticos de Lope en un formato de texto plano, para al menos ofrecer una alternativa y remediar el uso de los textos en línea deficientes y no fiables. Este capítulo constituye una aportación altamente erudita, informativa y crítica al terreno de la edición digital, que toma en consideración un amplio abanico de ejemplos y deja los cimientos para la creación de ediciones digitales críticas según los métodos fijados por la crítica textual.

Un segundo grupo de contribuciones centran su atención en la edición crítica de o en torno a un único autor. Mercedes Blanco (Sorbonne Université) y Aude Plagnard (Université Paul-Valéry) reflexionan sobre su experiencia en el proyecto *Íñólemos*, cuyo principal objetivo es la edición y el estudio de la obra completa de Luis de Góngora, así como de los documentos relativos a su recepción a lo largo de los siglos XVI y XVII. Entre los textos, incluyen comentarios sobre su obra, escritos a favor o en contra, o debates en torno a su obra, lo que se conoce en definitiva como la “polémica gongorina”. Su edición, colectiva, filológica y digital, disponible ya en línea, da lugar a una serie de reflexiones sobre la naturaleza de la edición filológica digital que nace justamente de su experiencia “empírica”, así como del rigor y sistematicidad de su metodología. Blanco y Plagnard trazan elocuentemente la línea que divide el paradigma tipográfico del paradigma digital y las diferentes razones que conducen a seguir uno u otro. Entre estas razones, se arguye la dimensión y diversidad de materiales del corpus a editar, cuestiones económicas, el carácter colaborativo, pero sobre todo la “entidad plural” del texto, que permite ofrecer más de una versión del mismo, e incluir paratextos de cualquier naturaleza, históricos o fruto de la crítica contemporánea, relacionados con la obra en un único sitio. Destaca este capítulo por su honestidad y claridad al evidenciar las dificultades de toda edición colaborativa y ofrecer las estrategias adoptadas para superarlas: la necesaria formalización de las diferentes secciones para crear un corpus homogéneo (por ejemplo, adopción por parte de todos los colaboradores del mismo patrón para la introducción) o el establecimiento de unos criterios ecdóticos. Explican también el

modelo de datos que contempla diferentes tipos de anotación: notas pertenecientes al mismo autor, notas de edición crítica que ofrecen un aparato positivo de variantes o señalan alguna variante relevante, y notas explicativas del editor. Altamente relevantes son las descripciones ofrecidas sobre los límites de la conversión de sus archivos originales en Word hacia el formato TEI, así como las consideraciones sobre las dificultades y los retos puestos por la codificación en XML-TEI. En fin, este capítulo ofrece un ejemplo elocuente sobre el potencial de reutilización de los datos textuales codificados para otras salidas, como por ejemplo las visualizaciones de las relaciones en la interlocución en una comedia entre dos o más personajes, o la intertextualidad en la polémica gongorina.

Eugenia Fosalba (Universitat de Girona) y Susanna Allés-Torrent (University of Miami) ofrecen una panorámica general del proyecto Pronapoli, concebido para el estudio de la estancia italiana de Garcilaso de la Vega que este poeta llevó a cabo de forma discontinua entre los años 1529 y 1536. La metodología que emplea Pronapoli en su inmersión total en el contexto italiano (histórico, cultural, literario) ha llevado a la creación de una plataforma digital donde se atomizan las distintas facetas de ese entorno inexcusable en las pesquisas de sus investigadores: una galería de autores relacionados con Garcilaso y su estancia en Nápoles, un geolocalizador de las academias que dan buena cuenta del ambiente intelectual del momento, además de una biblioteca digital con transcripciones y digitalizaciones de manuscritos poco conocidos y accesibles, así como bibliografías primarias y secundarias. La culminación de esta inmersión total, en una extensión de la plataforma que ha adquirido el título de *Soledad amena*, es la edición crítica y digital de las obras de corte horaciano, y a continuación, de la edición de la obra completa del gran poeta, que aparece inmersa en estas mismas coordenadas contextuales, gracias a un diseño integrador. Se espera que cuando la edición digital de las obras realizadas en el proyecto anterior se abra al público en general, el impacto de Pronapoli crecerá exponencialmente. En la actualidad, no existe una edición, ya sea impresa o digital, que cubra las cinco odas latinas. Tampoco hay una edición crítica de estas odas, y la que se ofrece, de acceso abierto, tiene la virtud de ser la más ampliamente anotada. Permite varios tipos de lectura: una más rápida, simplemente para leer el texto sin obstáculos (con meras notas explicativas), y otra más erudita, con excursos más extensos o comentarios sobre el aparato crítico, dirigida al lector académico, ya sea erudito o estudiante. El capítulo dedicado a este proyecto editorial insiste en la razón de una edición de estas características, y en la necesidad de crear un modelo de representación de datos que dé buena cuenta de la fuente primaria y del tipo de elementos que de esta se desprenden. Así también, reflexiona sobre el proceso editorial que se despliega en tres fases principales: el de la transcripción y establecimiento del texto, la codificación y la visualización o presentación.

Un último grupo con dos aportaciones, interesantísimas por sus puntos de conexión y divergencias, se centra en la creación de corpus y el uso de técnicas

de procesamiento de lenguaje natural para su anotación. Por un lado, Helena Bermúdez Sabel (JinnTec), Clara Martínez Cantón (UNED) y Pablo Ruiz Fabo (Université de Strasbourg) ofrecen las directrices esenciales que han llevado a la creación de DISCO (*Diachronic Spanish Sonnet Corpus*), un corpus que al momento contiene 4.530 sonetos escritos en español pertenecientes a diferentes regiones y épocas. El subcorpus para textos del Siglo de Oro se compone de 477 poetas y 1.088 sonetos. El corpus está anotado a partir de técnicas de procesamiento del lenguaje natural (PLN) y contiene informaciones métricas tales como el número de sílabas por verso, la acentuación, la rima y la existencia o no de encabalgamiento. Uno de los aspectos sumamente interesantes de esta contribución es la constatación de que los resultados de una anotación automática pueden detectar problemas de transmisión textual. Su experiencia demuestra, así, cómo la anotación automática basada en técnicas de PLN pueden ser integradas en los flujos de trabajo para múltiples fines, pero sobre todo, y en el caso de este monográfico, para el establecimiento, el análisis y la fijación del texto. Al detectarse errores textuales mediante técnicas automatizadas, los editores pueden centrar sus esfuerzos en la revisión manual de aquellos pasajes más problemáticos o que induzcan a un error (por ejemplo, la detección de un verso no rimado, error en el número de versos o sílabas, etc.). Además, esta contribución ofrece una iluminadora descripción del flujo de trabajo, su metodología y la lógica detrás del establecimiento de los criterios para la edición y la limpieza de los textos que otorgan coherencia al corpus.

Por otro lado, Borja Navarro Colorado (Universidad de Alicante), partiendo de su experiencia con el *Corpus de Sonetos del Siglo de Oro* desarrollado en el proyecto ADSO (*Análisis Distante del soneto castellano del Siglo de Oro*), propone el concepto de “edición digital enriquecida”, basada en un sistema de anotación multinivel de un corpus literario. El objetivo de este tipo de anotación es el de fijar un análisis a través de la integración de informaciones a diferentes niveles de descripción lingüística y literaria (como por ejemplo, métrica, sintáctica, semántica, temática, etc.). Además, Navarro propone una serie de requisitos inherentes a toda edición digital enriquecida: interoperabilidad (uso de estándares como XML-TEI y compatibilidad con otros corpus), perspectivismo (relación entre los diferentes niveles de análisis para mantener la unidad textual), unidad del corpus (consistencia), sencillez y claridad (posibilidad de edición manual del código). Al valorar estos requisitos, Navarro analiza también ventajas y desventajas, y evalúa el uso de herramientas y lenguajes de marcado. Algunos casos discutidos corresponden a la anotación de la información sintáctica o a la sobreposición de informaciones, donde el lenguaje XML-TEI pone problemas de verbosidad, en el primero, y de anidación en el segundo. En el caso del *Corpus de Sonetos del Siglo de Oro* hasta el momento el análisis se ha centrado en la detección del hipérbaton como rasgo estilístico. Para ello los textos se han marcado a tres niveles diferentes: métrico, léxico y sintáctico. Para los dos primeros tipos de anotación se ha utilizado XML-TEI, mientras que para la representación de la información sintáctica

se ha usado el formato CONLL. Esta contribución da buena cuenta de la complejidad de la anotación de un corpus a niveles diferentes con informaciones de naturaleza heterogénea y de los retos informáticos a tener en cuenta.

Como indican Sabel, Martínez y Ruiz, ambos corpus coinciden e incluyen algunos de los mismos textos (concretamente, 17 autores y 333 sonetos en común), la mayoría de los cuales han sido recabados de la Biblioteca Virtual Miguel de Cervantes. Ambas aportaciones son altamente iluminadoras para entender el procesamiento del lenguaje natural y qué aplicaciones prácticas tienen estas en los estudios literarios y para la crítica textual. Mientras que Sabel, Martínez y Ruiz ejemplifican lo que las técnicas automáticas de procesamiento de texto pueden aportar a la tradición textual y al establecimiento de un texto, Navarro añade complejidad a la anotación lingüística y semántica que puede conllevar una edición digital.

Como no podía ser de otra manera, en este monográfico, muchos proyectos relevantísimos relacionados con los textos del Siglo de Oro han quedado fuera, pero huelga decir que no ha sido nuestra intención ofrecer con exhaustividad la variada casuística de lo que se está llevando a cabo. El objetivo ha sido ofrecer una serie de contribuciones que dialogaran entre ellas, ya fuera de manera implícita o explícita, y que dieran buena cuenta de lo que constituye hoy en día la edición filológica de textos del Siglo de Oro.

Editar libros de caballerías en la era digital: la Biblioteca Digital del Proyecto Mambrino

Anna Bognolo

Università di Verona
ORCID: 0000-0002-1194-4520
anna.bognolo@univr.it

Stefano Bazzaco

Università di Verona
ORCID: 0000-0002-3496-3687
stefano.bazzaco@univr.it

Resumen

El presente trabajo trata la creación y publicación de la Biblioteca Digital del Proyecto Mambrino, grupo de investigación de la Universidad de Verona que se ocupa de los libros de caballerías italianos de inspiración española publicados en Venecia en los siglos XVI-XVII. Se aportan informaciones acerca de las obras del corpus, los objetivos del proyecto, y la realización de la Biblioteca Digital, cuyo centro recae en las ediciones digitales piloto del libro trece del Ciclo de Amadís de Gaula, el Sferamundi di Grecia de Mambrino Roseo da Fabriano. Finalmente, se aclaran los futuros avances e integraciones para sustentar la difusión y la accesibilidad a los recursos.¹

Palabras clave

Proyecto Mambrino; PRIN Mapping Chivalry; edición digital académica; accesibilidad; Amadis de Gaula.

Abstract

Editing Chivalric Romances in the Digital Age: the Case of Mambrino Project Digital Library.

1. El presente trabajo se inscribe en el marco del Progetto Mambrino (Gruppo di Ricerca sul Romanzo Cavalleresco Spagnolo), del PRIN *Mapping Chivalry. Spanish Romances of Chivalry from Renaissance to 21st century: A Digital Approach* (2017-2023), del proyecto de Excelencia “Inclusive Humanities 2023-2027” del Dipartimento di Lingue e Letterature Straniere de la Università di Verona.

This work deals with the creation and publication of the Digital Library of the Mambrino Project, research group at the University of Verona that deals with Italian Romances of Chivalry published in Venice in the 16th-17th centuries. Information is provided about the works in the corpus, the goals of the project, and the creation of the Digital Library, with digital editions of the thirteenth book of the Amadis of Gaul Cycle, the Sferamundi of Greece by Mambrino Roseo da Fabriano, at the center. Finally, future developments and integrations to support the dissemination and accessibility of resources are clarified.

Keywords

Mambrino Project; PRIN Mapping Chivalry; Digital Scholarly Edition; accessibility; Amadis of Gaul.

Libros de caballerías entre España e Italia: el corpus

En la primera edad moderna, España fue un foco de irradiación de nuevos tipos de ficción en prosa. Géneros como la novela sentimental, picaresca, celestinesca, pastoril y caballerisca salieron de España para propagarse de especial manera por Italia y Francia, y más tarde a toda Europa, para alimentar el mercado del libro en el primer siglo de la era de Gutenberg (Cabo Aseguinolaza 2013).

En la prosa castellana del siglo XVI, sobre la base del éxito del *Amadís de Gaula*, los libros de caballerías son un acervo de novedades y experimentaciones. Bajo la presión del interés comercial de la imprenta, en un contexto literario de hibridismo y maleabilidad, los libros de caballerías se apoderan de formas y contenidos de otros géneros y se hacen cada vez más interesantes (Marín Pina 2011).² Constituyen un patrimonio de formas y contenidos que se expande, y

2. Los libros de caballerías no son, como se ha creído, una forma literaria desdeñable por mal compuesta y aburrida (“todos una misma cosa” decía el canónigo del cap. I, 47 del *Quijote*). Para traer a colación un solo ejemplo recientemente publicado en edición moderna, *El Mexicano de la esperanza, caballero de la fe* (1583) de Miguel Daza, estudiado por Ana Martínez Muñoz, no es solo un libro de aventuras con episodios de novela pastoril o sentimental, sino una poliantea de erudición humanística con discusiones filosóficas, digresiones históricas y geográficas, excursos enciclopédicos y comentarios moralizantes. Es también una novela en clave que esconde persona-

encuentra soluciones técnicas que se mantienen hasta en las novelas de hoy (Brooks 2004; Pavel 2005; Mazzoni 2011).

El *Proyecto Mambrino* estudia precisamente la recepción de los libros de caballerías hispánicos en Italia y en Europa; los responsables científicos son Anna Bognolo y Stefano Neri, y el equipo de trabajo está formado por investigadores del Departamento de Lenguas y Literaturas Extranjeras de la Universidad de Verona (Stefano Bazzaco, Federica Zoppi) y de otras universidades italianas (Paola Bello-mi, Giovanni Cara, Claudia Demattè, Tiziana Mancinelli, entre otros).

El corpus de interés del proyecto está constituido por más de cincuenta obras en lengua italiana, traducciones del español y continuaciones de originales, que se difundieron en más de trescientas cincuenta ediciones. El grupo de investigación está censando todos los ejemplares de estas ediciones, conservados en las bibliotecas de todo el mundo, en una base de datos bibliográfica de la que se hablará más adelante, que contiene por el momento alrededor de mil seiscientas entradas.

A pesar de su importancia, el corpus caballeresco italiano de derivación española es poco conocido y ningún texto se ha publicado en edición moderna. La falta de dedicación de los estudiosos de literatura italiana se debe a la marginalidad del género respecto al canon nacional; a esto se añade la dificultad debida a la extensión de los textos, que hace poco rentable consagrarse a la tarea. Por consiguiente, poner los textos a disposición para que se conozcan ha sido un objetivo fundamental del proyecto, así como la publicación en una biblioteca digital pareció la solución más adecuada, porque permite optimizar la consulta de un corpus tan vasto y reducir los costes de edición. Permite además integrar en un único sitio varias novelas, de manera que una interrogación transversal y cruzada pueda revelar inmediatamente las continuidades o diferencias, e individuar recurrencias de personajes, lugares y aventuras dentro de la totalidad del mundo narrativo. Finalmente, la ventaja de editar los libros en la red es que se hacen prontamente accesibles a la comunidad internacional.

Historia del *Proyecto Mambrino*

El *Proyecto Mambrino* nació en 2003.³ Su nombre no se debe solamente al famoso yelmo conquistado por el paladín Rinaldo y reconquistado por Don Quijote, sino sobre todo al nombre del traductor y escritor Mambrino Roseo da Fabriano, el prolífico autor de casi todas las traducciones y continuaciones italianas de libros de caballerías españoles (Bognolo 2010, Bognolo y Neri 2014).

lidades de la alta nobleza, a la vez que plantea una interlocución irónica con el lector acerca del estatuto de la ficción, con juegos metatextuales de manera casi cervantina (Daza 2019).

3. Proyecto Mambrino, <<https://www.mambrino.it/it/>>.

El objetivo del proyecto fue ambicioso desde el comienzo: se proponía llevar a cabo una exploración de los libros italianos siguiendo las pautas de las guías de lectura de libros de caballerías castellanos que publicaba el Centro de Estudios Cervantinos de Alcalá de Henares. A partir de los ciclos de *Amadís de Gaula* y de *Palmerín de Olivia*, se inició un trabajo de equipo en colaboración con estudiantes de grado y de máster que redactaron cada uno un trabajo final dedicado sistemáticamente a un volumen del ciclo, con el análisis de cada texto, el resumen, el índice de nombres propios y el estudio de un tema. Al mismo tiempo se desarrollaron varias tesis doctorales dedicadas a textos clave de nuestro corpus.⁴

Una financiación conseguida en 2010 permitió llevar a cabo la digitalización masiva de ejemplares conservados en bibliotecas de la región de Véneto, como la biblioteca Cívica de Verona y la Biblioteca Bertoliana de Vicenza, que fueron en parte publicados en discos DVD.⁵ También se creó la primera página web que sigue abierta, donde publicamos las mismas colecciones digitalizadas y abrimos una sección dedicada a la vida y obra del escritor Mambrino Roseo de Fabriano.⁶

En 2013 publicamos el *Repertorio* del ciclo de *Amadís de Gaula* (Bognolo, Cara y Neri 2013) que representaba la coronación de todo el recorrido de las tesis antes recordadas. El repertorio contiene un estudio general sobre la vida del autor y el éxito de los libros de caballerías en Italia y en Europa, y un resumen y un índice de personajes de cada uno de los trece libros originales italianos. El análogo *Repertorio* del ciclo de *Palmerín de Olivia* está en preparación (Bellomi, Bognolo, Neri y Zoppi en prensa). El paso siguiente será justamente transportar a la red y hacer consultable en acceso abierto todo este abundante material: es decir, la edición digital interactiva que estamos preparando.

4. Las tesis doctorales son: la de Stefano Neri sobre arquitecturas maravillosas (Neri 2007a y 2007b), la de Linda Pellegrino con la edición del *Florisel de Niquea* (Silva 2015), la de Stefania Trujillo sobre el tema del doble en Feliciano de Silva (codirección entre Anna Bognolo y Alberto del Río, Universidad de Zaragoza), la de Stefano Bazzaco, con la edición del *Leandro el Bel* (Luján 2020). Se ha leído recientemente la tesis de Giada Blasut, que contiene la edición del *Silves de la Selva* (codirección entre Anna Bognolo y José Manuel Lucía Megías, Universidad Complutense de Madrid) y se está redactando la de Manuel Garrobo Peral sobre una base de datos de las fiestas en la serie de *Amadís de Gaula* (codirección entre Stefano Neri y Nieves Pena Sueiro, Universidad de Coruña y Proyecto BIDISO).

5. Los discos DVD contenían la reproducción en facsímiles de los ejemplares digitalizados acompañados por una ficha de descripción bibliográfica. En aquel momento pudimos llevar a cabo la transcripción de dos de los libros elegidos porque eran ediciones príncipes, *Il secondo libro delle prodezze di Splandiano* (1564) y *Aggiunta al Florisello (Le prodezze di don Florarlano)* (1564), que fueron las primeras ediciones consultables de libros de caballerías italianos (Bellomi, Bognolo, Colombini y Neri 2011-2012).

6. Proyecto Mambrino, *Vita di mambrino Roseo* <<https://www.mambrino.it/it/mambrino-roseo/vita-di-mambrino-roseo>>, *Le opere* <<https://www.mambrino.it/it/mambrino-roseo/le-opere>>.

Relación con otros investigadores y otros proyectos. *Mapping Chivalry*

Los seminarios de *Historias Fingidas*, que se inauguraron el 2010, nos permitieron reanudar la red de relaciones con estudiosos expertos de la materia y abrir las puertas a otros investigadores, mientras daban ocasión a los más jóvenes de percatarse de la riqueza cultural implicada en el estudio de los libros de caballerías. La convocatoria de los seminarios con periodicidad anual dio lugar a la fundación de la revista del mismo nombre *Historias Fingidas* (ISSN 2284-2667) donde se publicaban las contribuciones de estas reuniones. La revista fue apoyada desde su nacimiento por un comité científico de alto nivel, gracias al apoyo del cual, mientras los estudios en este campo se multiplicaban, llegamos a conseguir al cabo de varios años el nivel más alto de evaluación de las revistas en Italia (*Fascia A*).

Las relaciones más sólidas y fructíferas se instauraron con algunos grupos de investigación de la península ibérica como el grupo *Clarisel* de la Universidad de Zaragoza guiado por Juan Manuel Cacho Bleuca y la revista electrónica *Tirant* de València, dirigida por Rafael Beltrán. También es importante el intercambio científico y didáctico con la Universidad de La Coruña, donde el seminario SIELAE, con Sagrario López Poza y Nieves Pena Sueiro a la cabeza, publica la Biblioteca Digital BIDISO.⁷ Finalmente, la Universidad de Verona ha establecido un convenio de intercambio académico con el Seminario de Estudios sobre Narrativa Caballeresca de la Universidad Nacional Autónoma de México, dirigido por Axayácatl Campos García Rojas. Estos proyectos alcanzan amplios resultados publicando una gran cantidad de artículos científicos, archivos y textos fiables, a la vez que permiten unas reuniones sistemáticas de investigación. La esperanza es que, gracias a la Red Aracne, muchos de los recursos producidos por estos grupos puedan confluir en un repositorio interoperable.

El Ministerio Italiano de la Universidad (MUR) financió en 2019 el proyecto *Mapping Chivalry. Spanish Romances of Chivalry from Renaissance to 21st century: A Digital Approach* (PRIN 2017. Prot. 2017JA5XAR) (Bognolo 2023) en el que participa un equipo más amplio que incluye grupos de investigación de las universidades de Trento, de Roma La Sapienza y de Salerno. En este proyecto nacional, con el apoyo de la empresa informática Net7 de Pisa,⁸ se han creado otras tres bases de datos: la universidad de Trento ha desarrollado MeMoRam, base de datos de motivos caballerescos (dirs. Claudia Demattè y Giulia Tomasi);⁹ la de Salerno el repositorio de textos de teatro caballeresco (dir. Daniele Crivellari);¹⁰ la de Roma La Sapienza la base de datos de las reescrituras y

7. BIDISO, <<https://www.bidiso.es/>>.

8. Net7, <<https://www.netseven.it/>>.

9. MeMoRam, <memoram.mappingchivalry.dlls.univr.it>.

10. Teatro Caballeresco, <teatrocaballeresco.mappingchivalry.dlls.univr.it>.

recreaciones modernas de los libros de caballerías (dir. Elisabetta Sarmati).¹¹ El equipo de Verona (dir. Anna Bognolo y Stefano Neri) mantiene activo el Proyecto Mambrino, del cual la parte más ambiciosa es la biblioteca digital conectada con la base de datos bibliográfica y la base de datos de Personajes, Lugares y Motivos narrativos.¹²

También el Departamento de Lenguas y Literaturas Extranjeras de la Universidad de Verona fue financiado como Departamento de Excelencia en los años 2018-2022 y 2023-2027, lo que nos ha permitido contratar para el Departamento a dos profesores de informática y dos técnicos e integrar los recursos en un Centro Digital que preservará el sitio y los datos a largo plazo. Esto ofrece una solución viable a los problemas de sostenibilidad, que es uno de los aspectos cruciales a la hora de garantizar la longevidad y la supervivencia de este tipo de trabajos.

La Biblioteca Digital

La Biblioteca Digital del Proyecto Mambrino (de aquí en adelante BDPM) tiene el objetivo de publicar y dar visibilidad a estos textos del siglo XVI inéditos modernamente, que fueron muy leídos en su época y ahora están olvidados y ausentes de las historias de las literaturas española e italiana (Bognolo 2018; Bognolo y Bazzaco 2019).

En la decisión sobre el tipo de edición que queríamos llevar a cabo, hemos tenido en cuenta varios factores, en relación con la dimensión del corpus y las condiciones materiales de la escritura y de producción impresa en los talleres de Venecia. Ante todo, era evidente la utilidad de publicar el mayor número de ediciones en un tiempo razonablemente breve; en segundo lugar, las exigencias de urgencia y de practicidad nos sugerían mantener la conformidad con los criterios de los Libros de Rocinante de la Universidad de Alcalá, una transcripción levemente modernizada y depurada de errores de un ejemplar de la edición príncipe. En nuestro corpus el criterio del respeto de la voluntad del autor no es pertinente, puesto que las variantes se deben más a ajustes o descuidos de colaboradores de tipografía y a las incertidumbres de la lengua italiana durante la reforma de Bembo (Trovato 1991). En conclusión, hemos juzgado escasamente viable el esfuerzo de realizar ediciones críticas y asumido el riesgo de editar un ejemplar lo más completo y correcto posible de la edición príncipe de la obra.

Por otro lado, como se verá, los resultados de la BDPM merecen validación científica en cuanto ediciones académicas digitales. El texto de las obras está preparado y publicado siguiendo criterios y métodos rigurosos y documentados

11. AmadisSigloXX, <amadissigloxx.mappingchivalry.dlls.univr.it>.

12. Biblioteca Digital del Proyecto Mambrino, <mambrino.mappingchivalry.dlls.univr.it>.

de manera que el trabajo del editor sea verificable por parte del lector (Mancinelli y Pierazzo 2020: 9). Junto a ello, la preparación de cada edición es conforme a los principios de la filología digital, que suponen la superación de los límites del espacio bidimensional de la ‘página’ impresa, en un ambiente hipertextual que permite navegaciones y fruiciones diferentes según las exigencias de los usuarios y no se puede imprimir sin una pérdida de información (Sahle 2016).

Manejando físicamente cualquier ejemplar de nuestro corpus, cualquier estudioso se podrá dar cuenta de su extensión y de las dificultades que se encuentran en la digitalización y transcripción: se componen con una letra cursiva clara pero apretada y la encuadernación estrecha muchas veces no permite abrir las páginas para fotografiarlas adecuadamente, tanto que algunas bibliotecas contactadas se han negado a enviar reproducciones digitales.¹³

La importancia de los libros de caballerías

Creemos que la ficción en prosa anterior a Cervantes merece ser más conocida, no solamente por su valor en su época, sino para enriquecer la tradición de estudios que podríamos llamar ‘historia internacional de la novela’. A partir de ensayos clásicos como el de Ian Watt (2017), la línea de reflexión mayoritaria que se afirmó ante todo en los países anglófonos no tiene en cuenta las aportaciones hispánicas anteriores al *Quijote*, con la excepción del *Lazarillo de Tormes* y de la picaresca. El desconocimiento de la riquísima tradición novelesca renacentista española es casi total y relega estas obras a la irrelevancia, o a una porción de saber especialista aislado y cerrado en la historia de la literatura nacional. En este mismo contexto los libros de caballerías, a causa de su abundancia, de su extensión y de su mala fama, suelen ser los hermanos olvidados, uno de los géneros perdidos que señaló Alan Deyermond (1975). Hasta hace poco, solo algunas historias de la literatura les dedicaban un capítulo que les haga justicia.

No obstante, estas obras merecen mayor atención puesto que formaron el eje de la literatura de evasión de la primera edad moderna, que alimentó la industria y el mercado del libro, que facilitó la alfabetización y la democratización de la lectura, sobre todo en el caso de las mujeres, y provocó el paso de la lectura en voz alta a la lectura silenciosa, con los cambios antropológicos consiguientes (Loretelli 2010). Los libros de caballerías fueron el género que unió generaciones de lectores europeos en las emociones por los mismos personajes, amores y aventuras, superando las barreras nacionales, lingüísticas y religiosas del Renacimiento.

13. A la transcripción con herramientas de reconocimiento automatizado de textos (HTR) se han sumado dos niveles de revisión. Para las seis partes del libro de *Sferamundi di Grecia* (1558-1565), entre los pocos ejemplares conservados, elegimos dos de la BNE, uno de la Biblioteca Universitaria de Santiago de Compostela, y tres de la Biblioteca Braidense de Milán.

El Proyecto Mambrino quiere poner su grano de arena en el estudio de la novela moderna europea editando estos textos olvidados para que puedan leerse, conocerse, consultarse y siga adelante la investigación acerca de los mismos.

La configuración general de la Biblioteca Digital del Proyecto Mambrino

El objetivo de la Biblioteca Digital del Proyecto Mambrino, como se ha mencionado, es ofrecer una herramienta de estudio que permita combinar los resultados de investigación, las ediciones de los textos, el censo bibliográfico y el estudio de las obras, con particular atención hacia el desarrollo narrativo y la detección de motivos caballerescos. Por lo tanto, en línea general, la BDPM se concibió como un espacio informativo complejo, es decir una suerte de *knowledge site* (Robinson 2003) que permite al usuario navegar entre distintos campos de interés. Se trata de un archivo donde una variedad de recursos, integrados entre ellos, ofrecen una exploración del corpus textual desde distintas perspectivas, buscando la participación de tipologías de usuarios con intereses no necesariamente coincidentes.

La BDPM se constituye con tres elementos principales. En primer lugar, la base de datos bibliográfica de las ediciones y ejemplares de los libros de caballerías italianos, a la cual se añaden las ediciones modernas de las obras, elaboradas según criterios científicos y estándares técnicos acreditados, y diversas bases de datos que contienen lugares, personajes y motivos. Esta última ha conllevado una interpretación sobre los textos que se fundamenta en los principios ya trazados por especialistas de la materia. Se presentan a continuación estos tres recursos, con especial atención a las técnicas de integración de la edición con la base de datos bibliográfica y los índices para el estudio del contenido de las obras.

1. El censo de las obras: la base de datos bibliográfica

El censo de los ejemplares conservados de las novelas caballerescas italianas es una de las tareas que interesaron al Proyecto Mambrino desde su fundación en 2003. Las obras, pertenecientes en su mayoría a los ciclos italianos de *Amadis y Palmerino*, a los que se añaden otros ciclos menores y obras sueltas, son traducciones, continuaciones de originales y adaptaciones de las más famosas obras que se produjeron en la península ibérica y que circularon impresas por toda Europa a lo largo de los siglos XVI y XVII. La gran difusión de estos textos fue por lo tanto un fenómeno global, lo cual ocasionó que los numerosos ejemplares que se imprimieron acabaron por leerse y guardarse, y se conservan ahora en numerosas bibliotecas europeas y americanas. Nos referimos a más de cincuenta obras, que se difundieron en más de 350 ediciones. Gracias al trabajo de los

investigadores del proyecto se pudieron registrar los ejemplares de estas ediciones, que, hasta el momento, alcanzan el número aproximado de mil seiscientas, y sigue en aumento.

El censo se publicó a partir de 2008 en la web del Proyecto Mambrino.¹⁴ Sin embargo, las tecnologías informáticas en esa época eran bastante limitadas: los listados, organizados por título y fecha de las obras, se configuraron como páginas estáticas que no permitían búsquedas cruzadas.

Cuando dispusimos de recursos económicos suficientes, decidimos mejorar esta primera versión e integrar a nuestro nuevo proyecto de biblioteca digital una base de datos de tipo dinámico, que aprovechara sistemas de búsqueda transversales y que pudiera ser puesta al día constantemente, evitando el riesgo de duplicar fichas y facilitando la incorporación de nuevas informaciones. Desde mediados de 2019, trabajamos en el diseño y a la implementación de la base de datos bibliográfica. La estructura está dividida en tres niveles jerárquicos: obras, ediciones y ejemplares, cada uno compuesto por fichas que contienen varios campos relacionados entre sí y los distintos niveles.

The screenshot shows the 'Database Bibliografico' interface. At the top, there is a navigation bar with the logo 'PROGETTO MAMBRINO' and several menu items: 'Il Progetto', 'Database Bibliografico', 'Edizioni Digitali', 'Indici Semantici', and 'Produzione scientifica'. Below the navigation bar, the main heading is 'Database Bibliografico'. On the left side, there is a 'Filtra i risultati' section with a search input field and several filter categories: 'Lingua' (Italiano, Spagnolo), 'Soggetti' (poema cavalleresco, romanzo cavalleresco, romanzo), 'Cicli' (17), 'Autori' (49), and 'Traduttori'. On the right side, there is a list of search results. The first result is 'Amadis de Gaula' by Rodriguez de Montalvo, Garci, with 118 entries. The second result is 'I quattro libri di Amadis di Gaula (1-4)' by Rodriguez de Montalvo, Garci, with 82 entries. The third result is 'Aggiunta al quarto libro di Amadis di Gaula (A4)' by Riccio da Fabriano, Mambrino, with 111 entries. The fourth result is 'Las sergas de Esplandián' by Rodriguez de Montalvo, Garci, with 5 entries. The fifth result is 'Le prodezze di Splandiano (5)' by Rodriguez de Montalvo, Garci, with 5 entries. Each result includes the title, author, cycle, and subject.

Fig. 1: filtros y entradas de la base de datos bibliográfica.

Los campos de cada ficha coinciden con los filtros y opciones de búsqueda disponibles en la interfaz de usuario (Fig. 1), los cuales permiten extraer resultados a partir de la lengua, el sujeto y la tipología de las obras, la pertenencia a un

14. Proyecto Mambrino: Spagnole Romanzerie – Ciclo di Amadis di Gaula, <<https://www.mambrino.it/it/spagnole-romanzerie/ciclo-di-amadis-di-gaula>>.

determinado ciclo, los nombres del autor o del traductor implicados en la publicación. Al acceder a una obra en concreto se ofrecen los metadatos generales, correspondientes a los mismos campos de búsqueda; en la parte inferior de la pantalla, por otra parte, se muestran las ediciones de la obra, ordenadas cronológicamente y dispuestas en forma de enlaces, para acceder directamente al nivel sucesivo, el de una edición específica.

En el interior de la ficha de cada edición, se señala el título original, así como aparece en las fuentes impresas, los metadatos de la publicación y los ejemplares conocidos, destacando la eventual presencia de digitalizaciones o ediciones digitales relacionadas por medio de dos distintos iconos gráficos. Actualmente estamos trabajando en la incorporación de otras informaciones pertinentes, como el nombre del dedicatario de la edición y el enlace a catálogos en línea nacionales e internacionales, como los catálogos OPAC SBN (Catalogo collettivo delle biblioteche del Servizio Bibliotecario Nazionale) y USTC (Universal Short Title Catalogue).

1563 - Venezia - Tramezzino, Michele

Opera di riferimento	Aggiunta al quarto libro di Amadis di Gaula (A4)	
Titolo completo	Aggiunta al quarto libro dell'istoria di Amadis di Gaula, novamente ritrovata nella lingua spagnuola et ridotta nello idioma italiano per M. Mambrino Roseo da Fabriano	
Metadati	Luogo	Venezia
	Editore	Tramezzino, Michele
	Data	1563
	Descrizione	8"; cors. rom.; [16], 480 cc.; a-b8, A-308
	Dedicatario	Bevilacqua, Ercole
	Repertori	USTC: 804249 ID ICCU: IT\ICCU\CNCE\001394 Palau y Dulcet, A., Manual del librero, Barcelona, 1948, p. 299.

Fig. 2: ficha de edición de la *Aggiunta al quarto libro di Amadis di Gaula*, 1563.

Al tener acceso al nivel sucesivo, se ofrecen, para cada ejemplar, los metadatos asociados a la biblioteca, la descripción material del ejemplar, el *permalink* del repositorio online de la biblioteca y, en caso que los haya, los enlaces a las digitalizaciones. En el margen inferior, añadimos también un enlace a la edición digital del ejemplar concreto que publicamos en nuestra web, de modo que el acceso a la misma pueda darse a partir de los distintos niveles jerárquicos de la base de datos.

Con respecto a la precedente versión, los avances son significativos y se fundamentan en una mayor flexibilidad de búsqueda y de consulta, lo cual convierte la base de datos bibliográfica en punto de acceso fundamental a los textos

de la BDPM. Por un lado, la navegación por niveles determina la organización de los datos, que pueden extraerse de los filtros de búsqueda; por otro, los metadatos favorecen la exportación a otros repositorios en línea como la red de Excelencia ARACNE NODUS¹⁵ y Europeana en el formato Dublin Core según el protocolo OAI-PMH (*Open Archives Initiative's Protocol for Metadata Harvesting*).¹⁶ El próximo paso que se dará en el ámbito del nuevo Proyecto de Excelencia de la Universidad de Verona (2023-2027) será el de desarrollar una ontología semántica para describir de forma exhaustiva estos materiales bibliográficos, favoreciendo la distribución de los metadatos en formato digital como datos enlazados abiertos y la generación automática de redes de conocimiento relacionadas con el género literario del que nos ocupamos.

2. Las bases de datos semánticas: personajes, lugares y motivos caballerescos

Las tres bases de datos de personajes, lugares y motivos caballerescos son recursos fundamentales para examinar el contenido de las obras que componen la BDPM porque permiten detectar similitudes y diferencias entre los distintos libros. Con respecto a la creación de estos índices, contábamos ya con una labor previa, que llevó a la publicación de los dos repertorios en papel ya citados. En estos dos compendios, junto con unos estudios bibliográficos preliminares sobre la organización, catalogación y difusión de ambos ciclos caballerescos, se encuentran los resúmenes de cada obra de la serie y un apartado final con los nombres de los personajes y sus apariciones dentro de la saga. El trabajo llevado a cabo fue una tarea imprescindible para el estudio del contenido de las novelas caballerescas italianas; sin embargo, la necesidad de extender la investigación, incluyendo los nombres de lugares y los motivos caballerescos, no se tuvo en cuenta a causa de las limitaciones de la publicación impresa, porque la consulta de los dos tomos se habría convertido en una operación engorrosa. Por otra parte, el medio digital, por su intrínseca naturaleza, prometía responder a estas necesidades de análisis y catalogación.

En primer lugar, creamos dos bases de datos distintas para las entidades nominales (*named entities*) de personajes y lugares, cada una organizada por fichas. Al acceder a una ficha, el usuario puede encontrar los principales datos

15. Sobre la red ARACNE NODUS, véase Pena Sueiro y Saavedra Places (2019); Collantes Sánchez (2023).

16. Red ARACNE - Criterios, <<https://www.red-aracne.es/recursos/criterios#pie1>>. Estamos de momento trabajando en la extracción de los metadatos de cada ficha en el formato Dublin Core, con el fin de facilitar su inclusión en el portal de Red ARACNE NODUS. Se prevé que en un segundo momento los metadatos relativos a las ediciones se incluyan en el mismo fichero XML de los textos editados.

relacionados con la entidad consultada: en el caso de los personajes se señalan el ciclo de pertenencia, el nombre regularizado, el nombre español, los pseudónimos, el género y una breve historia del sujeto, que da cuenta de sus principales intervenciones dentro de cada libro del ciclo; para los lugares, por otro lado, se indican el nombre regularizado y la tipología, distinguiendo entre lugares reales e imaginarios.

Todos los datos incluidos en las fichas, que podrán extenderse en relación con las necesidades del proyecto, permiten realizar una búsqueda cruzada entre las distintas entidades nominales. Por ejemplo, si el usuario tiene la intención de investigar las doncellas guerreras, se pueden extraer todos los personajes femeninos que aparecen dentro de un ciclo; y del mismo modo, si quiere aplicar herramientas de georreferenciación para ver la distribución espacial de una obra en concreto, todos los lugares reales citados.

Con respecto a los motivos caballerescos, por otra parte, hemos previsto que, a través del portal principal, el usuario pueda realizar sus propias búsquedas por palabras clave, aplicar diferentes filtros y acceder a categorías de motivos divididos en tres apartados principales: motivos de corte, de aventura y de lo maravilloso, a los cuales se añaden las secciones de modificadores (es decir de elementos que califican un motivo sin suponer un cambio determinante en su disposición en el entramado narrativo) y de motivos metanarrativos (Zoppi 2022; Bognolo 2023). Dentro de cada sección se encuentran las fichas de motivos, que no son exclusivas de ningún apartado: en cada ficha se indica el nombre del motivo en tres idiomas (italiano, español e inglés), un código de referencia derivado de otros índices conocidos (por ejemplo los de Thompson, Rotunda y Guerreau-Jalabert), una breve descripción y una bibliografía de referencia. Además, se señalan las conexiones con otros índices de motivos en papel y en línea, para resaltar los aspectos de continuidad y las innovaciones del género caballeresco con respecto a otras tradiciones narrativas.

Finalmente, dentro de cada ficha contenida en las tres bases de datos, se incluyó una sección de ocurrencias, es decir de las apariciones de tal personaje, lugar o motivo en las obras que componen la biblioteca digital. Esta solución constituye otra vía de acceso a las ediciones digitales del proyecto, consiente la exploración de los textos e impulsa la navegación y la comparación entre episodios.

La base de datos es de tipo relacional y tiene la peculiaridad de memorizar la información en los ficheros XML-TEI de los resúmenes de las obras: de este modo, el portal se alimenta de manera automática, porque cada ocurrencia se genera a partir del procesamiento de las informaciones que se insertan a nivel de *backend*. Esto permite que al incorporar una nueva edición digital y al actualizar las bases de datos, las nuevas entradas vayan incorporándose a los ficheros TEI que se suben a una database XML nativa.¹⁷ Esta solución implica que con la puesta en línea de

17. Para ello, se ha utilizado Exist <https://exist-db.org/exist/apps/homepage/index.html>.

nuevos contenidos textuales se potencian las posibilidades de interacción de los usuarios, porque tendrán no solamente unas bases de datos cada vez más extensas a su disposición, sino también unos resúmenes enriquecidos.

3. La edición digital académica del corpus caballeresco italiano

Como ya hemos subrayado, el núcleo de la Biblioteca Digital son las ediciones digitales de las obras del corpus, que es una forma de presentación de los datos. En general, nuestra herramienta presenta algunas características fundamentales de los primeros prototipos desde el punto de vista de la producción y presentación de los contenidos (por ejemplo, la vista en paralelo de digitalizaciones y texto, o la producción de documentos XML-TEI). Esta continuidad de modelos y estándares es la base que asegura cierta estabilidad metodológica en las prácticas editoriales digitales (Pierazzo y Del Turco 2023: 115) y acrecienta la usabilidad por parte del usuario.

Guiados por la intención de ofrecerles a los estudiosos una herramienta dinámica, que permitiera integrar el texto de las obras con la base de datos bibliográfica y los índices semánticos, concebimos una visualización basada en una versión personalizada del software de publicación TEI Publisher.¹⁸ Nos situamos, por lo tanto, en el campo de las ediciones *prêt à porter*, es decir, las publicaciones que aprovechan softwares de visualización ya existentes, adaptándolos a las necesidades específicas de un proyecto en concreto (Pierazzo 2016).

Nuestra edición aparece como una página web tripartita en que se integran las imágenes digitalizadas de la fuente con las transcripciones y los resúmenes capítulo por capítulo, que constituyen la capa de estudio semántico del texto. Para generar estos tres recursos, elaboramos un flujo de trabajo de tipo colaborativo que comentamos a continuación.

3.1. Digitalización de las fuentes y relaciones con bibliotecas

La foto-reproducción de los ejemplares del corpus caballeresco italiano fue una tarea del Proyecto Mambrino desde su fundación. Remontan al año 2010 las primeras iniciativas de digitalización con bibliotecas locales, como la Cívica de Verona y la Bertoliana de Vicenza, pero esta perspectiva de llevar a cabo el trabajo con medios propios cedió el paso a iniciativas colaborativas de más amplio alcance. Dentro de estas, constituyó un momento fundamental la propuesta de cooperación con la Biblioteca Nacional de España (BNE), puesto que esta institución se mostró muy abierta a una colaboración con el proyecto. Al respecto,

18. Tei Publisher, <<https://teipublisher.com/index.html>>.

aclaramos que la BNE representa un ejemplo ejemplar de gestión del patrimonio cultural, ya que en estos años se está acercando a los mismos recursos de preservación y distribución sugeridos por los mayores planes europeos, y ofreciendo como resultado excelentes prácticas de producción, utilización y difusión de las imágenes digitalizadas.¹⁹

En nuestro caso, la BNE se ocupó de digitalizar las ediciones príncipes de obras de nuestro interés, a partir del corpus piloto que escogimos, es decir, el decimotercer libro del ciclo de Amadís, titulado *Sferamundi di Grecia*, que es una obra original en seis tomos de la pluma del polígrafo veneciano Mambrino Roseo da Fabriano. A continuación, gracias a la financiación del proyecto de excelencia, pudimos hacer públicas las digitalizaciones de la BNE gracias a unos servidores IIIF (International Image Interoperability Framework) del Departamento de Lenguas y Literaturas Extranjeras de la Universidad de Verona, favoreciendo la posibilidad futura de una difusión a largo plazo que pudiera alimentar a la vez la visibilidad del proyecto y la de la biblioteca. El IIIF es un formato nacido en el contexto de la biblioteconomía que se fundamenta en la interoperabilidad y la difusión de foto-reproducciones de la fuente como un conjunto de especificaciones estándar de API según una estructura Json-LD (Linked Data). Emplear tal modelo descriptivo para nuestras imágenes acrecienta de forma notable la accesibilidad, fruición y preservación del facsímil digital (Mancinelli 2021: 74-81). Las imágenes se presentan como piezas de un mosaico que la arquitectura de la edición comulga en un único artefacto digital, pero que a la vez son exportables como objetos separados, en el respeto de los principios FAIR, sobre todo por lo que atañe al procesamiento de metadatos y a la reutilización de las imágenes en otros proyectos. En este sentido, la BNE podrá aprovechar en el futuro nuestro trabajo de sistematización de los materiales archivados y favorecer un uso extensivo de las mismas digitalizaciones en red.

Esperamos que esta primera experiencia sea el punto de arranque para nuevas colaboraciones que no se limiten a la adquisición de imágenes, que son planes de menor interés desde el punto de vista de la investigación porque limitan la difusión y reutilización del patrimonio documental digitalizado. Sin embargo, nuestra experiencia al respecto nos ha permitido constatar que todavía existe

19. La página web de la BNE resulta de sumo interés al respecto. Ya en años anteriores a la pandemia Covid-19 la institución había difundido la guía “Proceso de digitalización en la Biblioteca Nacional de España” <https://www.bne.es/sites/default/files/repositorio-archivos/ProcesoDigitalizacionBNE_0.pdf>, que sigue representando un ejemplo encomiable de documento para la gestión de proyectos de digitalización a nivel europeo. Sin embargo, aún de más interés son las recientes noticias acerca de nuevos planes para la preservación digital y la reutilización de los artefactos digitales derivados de sus fondos documentales (1 marzo 2023, <<https://www.bne.es/es/noticias/nuevo-impulso-digitalizacion-acceso-reutilizacion-digital-bne>>). Sobre la integración de los metadatos de imágenes digitalizadas al formato de datos enlazados se consulten Santos Muñoz, Molleja López y Vizcaíno Ruiz (2022) y el volumen editado por Spadini, Tomasi y Vogeler (2021), punto de referencia para la comunidad internacional.

cierta reticencia al diálogo sobre la interoperabilidad por parte de algunas instituciones. El deseo es que esta situación pueda cambiar cuando se fijen de forma contundente los planes de digitalización previstos por la agenda europea,²⁰ que de momento parecen más bien unas sugerencias que necesitan cierto tiempo para arraigar de forma definitiva.

3.2. *Transcripción semi-automática de las fuentes*

A partir de las digitalizaciones, nos interesaba transcribir los textos. Como se ha dicho, puesto que el corpus es muy amplio, la transcripción manual no parecía rentable: exploramos por lo tanto la posibilidad de emplear herramientas de reconocimiento de textos para los impresos antiguos (HTR – Handwritten Text Recognition) que permitieran extraer transcripciones de forma no supervisada, las cuales después se corregirían por medio de un atento proceso de revisión.

Tras una valoración de los softwares de HTR disponibles, decidimos emplear la plataforma Transkribus, una herramienta elaborada por el grupo DEA (Digitalisierung & Elektronische Archivierung) de la Universidad de Innsbruck, en el contexto del plan Horizonte2020, actualmente gestionada por la Cooperativa Europea READ (Retrieval and Enrichment of Archival Documents) (Muehlberger *et al.*, 2019).

La plataforma fue entrenada para el reconocimiento de la letra cursiva del XVI por medio de la transcripción manual de algunas páginas de las fuentes y, pasando por la elaboración de unos específicos modelos de HTR, permitió transcribir de forma semi-automática las obras piloto del corpus con un margen de error mínimo que se acercaba al 1%, dependiendo de la calidad de las digitalizaciones empleadas (Bazzaco 2018). Estas transcripciones en bruto constituyeron el primer paso para la fijación del texto de las obras, porque de entrada se revisaron por grupos de estudiantes y en un segundo momento se editaron gracias a la intervención de especialistas, colaboradores del proyecto.

El empleo de herramientas de transcripción automática representa un claro aspecto de interés del proyecto desde el punto de vista metodológico, porque, a la luz de sus aplicaciones en otros ámbitos, coincidió con un primer momento de transferencia de conocimiento hacia una comunidad de usuarios interesados. Es el caso, por ejemplo, de los estudiantes, a quienes se formó en la producción automática del texto y su revisión. Para este fin, en los últimos cuatro años se impulsó la investigación sobre reconocimiento de textos por medio de varios talleres y seminarios, insistiendo en la utilización de Transkribus como una herramienta capaz no solamente de aligerar la labor de transcripción, lo cual res-

20. Digital Agenda for Europe, <<https://www.europarl.europa.eu/factsheets/en/sheet/64/digital-agenda-for-europe>>.

ponde a una función eminentemente instrumental, sino también como propulsora de un cambio en las prácticas editoriales, porque constituye de por sí un instrumento capaz de potenciar de modo determinante el flujo de trabajo ecdótico tradicional (Bazzaco *et al.* 2022).

En nuestro caso, a causa de la extensión del corpus, seguimos la idea, bien trazada por Sahle (2016) y retomada por Alvite-Rojas Castro (2022), de ofrecer una edición científica porque basada en la captura de “algunas propiedades de la fuente originaria”, lo cual conlleva que sea crítica en el sentido de que es “el resultado de la reflexión y del análisis de las fuentes textuales” (2022: 3).²¹ Esperamos que nuestras experimentaciones en el área de la transcripción automática de impresos antiguos permitan el avance de otros proyectos, poniendo las bases para una participación más activa de los filólogos en el proceso de migración de los textos al espacio del World Wide Web.²²

3.3. El modelo de representación de los datos: la codificación a la luz de la integración de recursos

Como es bien sabido, la modelización representa un aspecto central en la formalización de los objetos analógicos y constituye un pasaje decisivo para la confección de una edición digital académica (Ciula y Marras, 2023). Esto es verdad por lo menos por dos razones. En primer lugar porque, a pesar de que contamos con muchos ejemplos de ediciones digitales de altísimo nivel científico y metodológico, falta todavía un modelo estándar, sobre todo por lo que atañe a la presentación de los datos, si bien la TEI sigue siendo el punto de referencia (Spence, 2014; Allés-Torrent, 2017: 5 y ss.; Alvite y Rojas Castro, 2022: 5). En segundo lugar, y consecuentemente, porque es la modelización, entendida como configuración conceptual de la edición en cuanto ambiente informativo para las estrategias de construcción de significado, la que guía las posibilidades de “empleo” por parte de sus usuarios.

Con respecto a esta cuestión, son significativas las palabras de Baillot y Busch, quienes sostienen que “la comprensión del entorno editorial por parte del lector está en la base de todas las decisiones editoriales que implica el desarrollo de la edición” (2021: 175, trad. mía). Así, la modelización, contextualizada dentro de un marco eminentemente semiótico, se percibe como momento primordial de cada edición digital académica, porque implica la definición de

21. Sobre este tema se vea también Allés-Torrent (2020: 69).

22. La plataforma Transkribus propone planes para usuarios individuales, de uso gratuito y limitado a un número de entrenamientos y reconocimientos mensuales; para usuarios académicos depende del plan escogido. Más información en: <<https://www.transkribus.org/plans>>. En el caso del Proyecto Mambrino, hemos elaborado un modelo de HTR para el reconocimiento de la tipografía cursiva, que pondremos a disposición de los usuarios en 2024.

un modelo abstracto y formal sobre el cual, después, hay que fundamentar la memorización de los datos (analógicos) y la selección de la tecnología de publicación (Mancinelli 2021: 69).²³

Dos principios esenciales son los que guiaron la preparación de las ediciones digitales del Proyecto Mambrino: primero, integrar un aparato que permitiera el estudio de los contenidos de las obras; segundo, ofrecer a los usuarios distintas modalidades de lectura y acceso al texto. Se trata, en otras palabras, de buscar una forma de navegación de las obras que respete la correspondencia texto-imagen, pero que integre también los resúmenes y unifique los recursos principales en una única pantalla. Pensamos en una infraestructura que permitiera conectar los materiales entre ellos y que respetara estándares de economía de la información, con el fin de no distraer la atención del usuario. De este modo, los metadatos responden a la voluntad de enlazar las ediciones con repositorios externos; mientras que las bases de datos permitirían engarzar los contenidos con datos y proyectos afines al ámbito de estudio.

Para cumplir con estos propósitos se procedió con la elaboración de un sistema de marcado que consintiera la navegación simultánea del texto y del resumen de la obra, siendo este último el canal de acceso a las bases de datos semánticas. Se pasó entonces a la creación de dos ficheros XML-TEI distintos.

El texto de la edición (fichero “source”)

El primer fichero XML-TEI a la base de nuestra edición contiene el texto de la transcripción revisado y reproduce la estructura lógico-material de la fuente impresa de la obra. Este fichero (que llamamos *source*) está compuesto por una selección (*subset*) de elementos relevantes para las finalidades del proyecto, que se incluyeron en el esquema de validación Relax-NG del Proyecto Mambrino generado con la herramienta ROMA.²⁴ En conformidad con las normas de la TEI P5, se previeron dos secciones principales: el encabezado TEI, que contiene los metadatos de la edición digital e informaciones sobre la fuente empleada, y el texto editado.²⁵

- El encabezado (*teiHeader*): está compuesto por un primer bloque de metadatos, insertados en el elemento `<fileDesc>` y sus principales subsecciones, que son por la mayoría obligatorias según las prescripciones del modelado básico de la TEI. En primer lugar, el elemento `<titleSt-`

23. Dicho en otras palabras, la modelización “implica la traduzione di schemi concettuali, di sistemi complessi di conoscenza e relativi oggetti, in modelli processabili computazionalmente o in paradigmi operazionali” (Ciula y Marras 2023: 51), lo cual se fundamenta en una interacción mutua entre estudios humanísticos y computacionales .

24. TEI: ROMA, <<https://roma.tei-c.org/>>.

25. Los ficheros XML-TEI son descargables desde el visualizador TEI Publisher de cada edición.

mt>, en que se indican título de la edición digital, autor y editor de la obra original y responsables de la edición, distinguiendo entre curadores, revisores y codificadores, cada uno relacionado con su propio código ORCID. Sigue pues el apartado <*publicationStmnt*>, que contiene metadatos sobre institución editora (el Departamento de Lenguas Extranjeras de la Universidad de Verona), lugar y fecha de publicación, licencia de distribución (normalmente CC BY-NC-ND 4.0, que no permite la distribución comercial y de obras derivadas). Además, para la identificación de la edición dentro de la BDPM, se añadió el elemento <*seriesStmnt*>, que informa sobre la colección editorial y el ciclo. El elemento <*sourceDesc*>, finalmente, restituye informaciones acerca del ejemplar editado, como la localización (*@type: catalogue_number*) y el enlace al registro de la biblioteca. Este conjunto de metadatos es de gran importancia porque, al ser estable y compartido entre todas las ediciones, no implica solamente inventariar la edición digital dentro del corpus, sino también asegurar la interoperabilidad con otros proyectos y la extracción automática de informaciones por medio de *parsers*. El nivel de detalle, sin embargo, puede extenderse por medio de otros metadatos opcionales: de momento estamos trabajando en la definición de un esquema de metadatos Dublin Core para facilitar la futura exportación al formato de datos OAI-PMH.

- El texto de la obra (*text*): desde el punto de vista de la estructura lógica, en el cuerpo del texto se identifican las divisiones textuales (elemento <*div*> combinado con atributos *@type*), el título y el cuerpo de cada capítulo (elementos <*head*> y <*p*>). A partir de esta estructura, se añadieron los elementos que explicitan los aspectos materiales del documento, como rupturas de páginas y de líneas (<*pb*> y <*lb*>). El marcado así concebido, que al parecer resulta muy básico, tiene el fin de proporcionar un texto desprovisto de una interpretación fuerte, que sea lo más objetivo posible. De hecho, se preservan solamente los elementos que son esenciales para los propósitos del proyecto: los saltos de página y de línea implican una visualización paralela del texto y de las digitalizaciones (obviando la necesidad de presentar también una versión diplomática del texto); los elementos que describen la estructura lógica, como los títulos de capítulos, posibilitan aplicarles una presentación con un estilo específico a distintas porciones del texto. El resultado es un documento que evita la verbosidad y redundancia del lenguaje de la TEI, fomentando la preservación y la reutilización del texto limpio por parte de otros usuarios.

Un aspecto de interés dentro del proyecto supone un último recurso que permite la navegación simultánea de texto y su resumen. Se trata de un identificador *xml:id* progresivo que situamos a la altura de las divisiones textuales de cada capítulo: este atributo pone en relación el título del capítulo con la porción de

resumen correspondiente. De este modo, se pudo eludir el notorio problema de las jerarquías sobrepuestas, con el resultado de que toda la base de estudio semántico de la obra queda reflejada en otro fichero TEI, anclado al texto y ordenado según la misma estructura de niveles anidados.

Los resúmenes de la obra (fichero “commentary”)

Como se dijo, al fichero textual de la obra se añade un segundo documento XML-TEI que contiene los resúmenes capítulo por capítulo de la misma, en el cual se señalan las entidades nominales (nombres de personajes y lugares) y los motivos caballerescos. Se trata en otras palabras de adoptar una perspectiva modular de tipo *stand-off*, donde se le delega a los resúmenes la función de reproducir parte del contenido semántico, mientras que el texto de las ediciones queda intacto y preparado para futuras integraciones. La idea es de concebir un entorno digital hecho por distintos componentes que son extensibles y reemplazable sin afectar la apariencia general de la plataforma, que pueda configurar la edición como un *growing work site* (Schillingsburg, 2009: párrafo 4 y ss.).

Este segundo documento es más simple que el precedente desde el punto de vista de la estructura:

- El encabezado TEI del fichero es menos extenso y es sencillo en su organización. En las secciones `<titleStmt>`, `<publicationStmt>` y `<seriesStmt>` se insertan las mismas informaciones que en el documento TEI relativo al texto, como el título, los responsables de la edición, la colección y el ciclo de pertenencia. Varía solamente el elemento `<sourceDesc>`, que no remite a la *cinquecentina*, sino al correspondiente repertorio en papel del que extraemos los resúmenes.
- El texto del documento está compuesto por dos secciones. La división inicial, que llamamos paratexto digital, contiene informaciones acerca de la obra impresa e incluye los nombres de autor, editor, dedicatario: cada uno de ellos está incluido en un elemento `<persName>` y presenta enlaces a repositorios de autoridades como VIAF (Virtual International Authority File) o DBpedia con fines de desambiguación. Sigue pues el cuerpo de los resúmenes, dividido por capítulos, en que se indican las entidades nominales y los motivos. En concreto, para las entidades insertamos el elemento `<name>` con sus atributos `@person` y `@place` en correspondencia de la primera aparición de una entidad en el capítulo; mientras que, con respecto a los motivos, conectamos el grupo de motivos que aparecen en un capítulo con el título correspondiente del resumen por medio del elemento `<interpGroup>`, de modo que el investigador al consultar los resúmenes en el visualizador de la edición tenga de entrada toda la información necesaria para el estudio de la obra en cuestión.

Esta solución es particularmente eficaz porque constituye la base para la alimentación de la plataforma, puesto que a cada inserción de una nueva edición en la BD se transmiten informaciones a las bases de datos semánticas, que incorporan las nuevas entidades y señalan el lugar de aparición dentro de la obra de forma automatizada. De tal manera, facilitamos la expansión de la plataforma y aseguramos una mayor participación por parte de colaboradores externos al proyecto, que necesitarán tan sólo conocer los principios básicos del etiquetado y confeccionar un buen documento TEI para implementar la plataforma.

3.4. Ediciones de lectura: los eBooks del Proyecto Mambrino

Con la intención de proveer al usuario una modalidad de acceso al texto distinta, pensamos en la creación de unos eBooks descargables que se caracterizaran como “ediciones de lectura”, porque suponen la presencia de materiales de soporte en forma de vínculos hipertextuales. Gino Roncaglia distinguió entre dos distintas modalidades de consulta de los textos (2009: 14 y ss.). La primera de ellas es llamada lectura *lean forward* y corresponde a un acercamiento activo al texto, concebido para recabar información, que seleccionamos y elaboramos: adoptamos esta postura por ejemplo cuando estudiamos o estamos ante la pantalla del ordenador, porque buscamos y procesamos la información que nos interesa. Por otra parte, la segunda modalidad es llamada lectura *lean back*, es decir tumbada hacia atrás, como cuando leemos una novela o vemos una película relajados en un sofá y dejamos que la información impulse nuestra imaginación sin que se necesite una intervención concreta dentro del proceso comunicativo. Como se puede apreciar, el libro impreso es a la vez una herramienta de investigación y un objeto ergonómico apto para la lectura relajada, siendo por lo tanto idóneo para ambas modalidades. Sin embargo, en la migración de un texto al entorno digital no se mantiene esta doble configuración y es prerrogativa del editor planear distintas modalidades de representación de los datos textuales.

En el caso de nuestra edición digital, la visualización web requiere una postura del tipo *lean forward*, lo cual la caracteriza más bien como un “objeto para usar” que un “texto para leer”.²⁶ Los eBooks, al contrario, se ajustan a una postura de tipo *lean back* porque proponen una lectura lineal del contenido textual y se adaptan también a situaciones de lectura en movilidad. De momento estamos trabajando en la creación de unos eBooks de los tomos del *Sferamundi* que cumplan con este propósito, pero que mantengan las informaciones semánticas, por ejemplo incluyendo unos índices finales de resúmenes, personajes, lugares y motivos con enlaces hipertextuales que remitan al texto de la obra.

26. Se vean al respecto: Pierazzo (2015: 147-51); Stinne Greve Rasmussen (2016: 119 y ss.); Allés-Torrent (2017: 15).

3.5. *Diseño de la edición*

Una vez preparados los facsímiles, las transcripciones, los resúmenes y los eBooks, el paso siguiente consistió en concebir un ambiente web que permitiera conjugar y manejar todos estos materiales. Por un lado, estábamos convencidos de que la plataforma de visualización tenía que agrupar todos los recursos en una única pantalla, evitando dispersar la atención del usuario en un constante vaivén entre los distintos componentes de la edición. Por otro, era necesario presentar los materiales de manera eficaz, para que la arquitectura web no resultara de difícil comprensión y se acomodara a las expectativas de lectura e interpretación de los usuarios.

Soportados por la empresa Net7 de Pisa concebimos una versión personalizada de la herramienta TEI Publisher que tuviera en cuenta ambas necesidades.

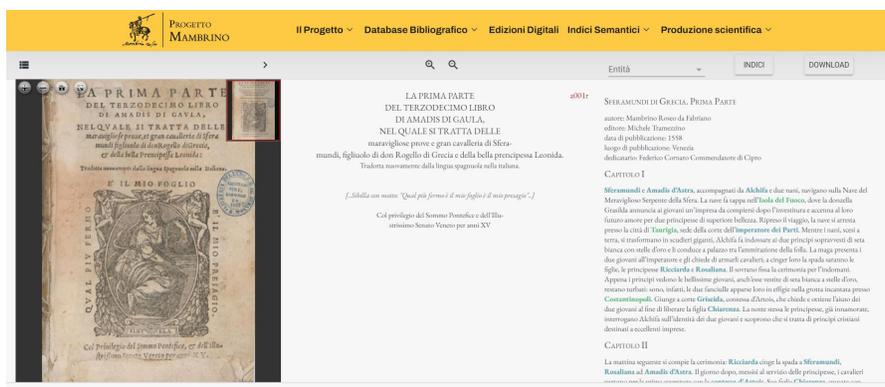


Fig. 3: diseño de la edición digital académica de la primera entrega de *Sferamundi di Grecia* (1558).

La presentación web, como se puede apreciar, se basa en el principio general de economía de la información y se ajusta al modelo de edición sinóptica ya empleado por otras ediciones digitales. En la columna de izquierda aparece la imagen facsímil en formato IIIF, acompañada por funciones de zoom y visualización a toda pantalla, mientras que en la columna central presentamos el texto transcrito de la obra. A estas dos secciones añadimos una columna a la derecha que contiene los resúmenes, cuya prerrogativa es la conexión del texto con las bases de datos semánticas.

Toda una serie de botones del visualizador permite manejar los tres contenidos principales de la edición: hay flechas para la navegación sincrónica de las páginas digitalizadas y del texto, mandos con funciones de zoom para el texto, un desplegable para la navegación de capítulos. En la sección del resumen, además, añadimos una función para activar la detección de entidades, que aparecen en distintos colores: al seleccionarla, el usuario accede a la correspondiente voz

del índice semántico, que se abre en una nueva pestaña. Por medio de estos mandos aseguramos que el texto de la edición siga siendo central en el espacio de interacción del usuario, pero también ofrecemos el acceso a otros recursos.

Además, incorporamos al diseño del visualizador un botón de *download*, que permite al usuario descargar y consultar en local un fichero comprimido que contiene los documentos XML-TEI del texto y del resumen, la versión eBook de la obra y un fichero TXT para el procesamiento con herramientas de lectura distante, para el estudio cuantitativo y el procesamiento automático de la información.²⁷

En la parte inferior de la página, justo debajo del visualizador, hay el apartado de metadatos. En él aparecen distintas informaciones: el responsable de la edición, la fecha de publicación, la licencia de distribución y los códigos de referencia del artefacto digital, como el identificador ISBN y el DOI. Pero también se incluyen enlaces a los distintos niveles de la base de datos bibliográfica, que permite volver a las fichas de obra, edición y ejemplar correspondientes al texto consultado. La página web de la edición, también en el caso de los metadatos bibliográficos, admite que el usuario pueda acceder a otras secciones de la BDPM, sin que por ello abandone la página principal del visualizador. De hecho, la ficha de la base de datos bibliográfica aparece en una nueva pestaña, lo cual asegura que siempre se pueda volver al texto de la edición en correspondencia del punto en que se había dejado la lectura. Se trata, en otras palabras, de volver a poner al centro de la edición el texto, evitando que la navegación resulte demasiado compleja y, por consiguiente, opaca para el usuario (Van Mierlo 2022: 117 y 118).

Conclusiones: difusión, sostenibilidad y accesibilidad

La Biblioteca Digital del Proyecto Mambrino estará disponible en la web desde enero de 2024. Como ya se ha mencionado, se trata de uno de los cuatro portales que constituyen el proyecto *Mapping Chivalry. Spanish Romances of Chivalry from Renaissance to 21st century: A Digital Approach*, dedicado al estudio de la literatura caballerescas ibérica y sus múltiples manifestaciones literarias. En concreto, la plataforma realizada por el grupo de investigación veronés se ocupa del censo, el estudio y la publicación digital de las novelas de caballerías italianas de derivación española del Renacimiento. Para responder a estos objetivos, se crearon distintas herramientas que permitieran una renovación y mejor difusión de los resultados de investigación del Proyecto Mambrino: una base de datos bi-

27. Pueden consultarse los trabajos de Franco Moretti producidos en el Stanford Literary Lab, por ejemplo: <<https://litlab.stanford.edu/projects/operationalizing/>> (cons. 19/01/2024). Sobre la estilometría y los estudios computacionales de atribución en relación con las literaturas ibéricas, véase la bibliografía citada en Bazzaco (2021).

bliográfica que registra los ejemplares conservados de las obras caballerescas, unas bases de datos semánticas que acrecientan las posibilidades de estudio del corpus, unas ediciones digitales académicas que faciliten el acceso a los textos. A pesar de que una parte del trabajo inicial de ideación e inicialización del portal está hecha, quedan todavía algunas cuestiones abiertas, sobre todo en lo que atañe a la preservación, difusión y utilización de los recursos generados.

Entre los objetivos futuros, quizás el más central es el tema de la difusión de la información, que tiene que fundarse en la integración con otros repositorios en línea. Con tal fin, se están dando varios pasos para la incorporación de los contenidos de la plataforma a varios repositorios, de manera que las ediciones sean accesibles a través de distintos portales. En primer lugar, por medio de la incorporación de un metabuscador de dominio en la página del proyecto Mapping Chivalry, los productos digitales de la BDPM serán puestos en relación con los tres portales elaborados por los grupos de investigación con que colaboramos, fomentando la agrupación de los contenidos y la conexión con otros proyectos del mismo ámbito. En segundo lugar, gracias a la adaptación de los metadatos al protocolo de distribución OAI-PMH, se podrán incorporar los productos de nuestra investigación a repositorios de más amplio alcance, como RED ARACNE y Europeana. La intención futura es que, gracias a la migración de los contenidos al formato de datos enlazados abiertos, nuestra biblioteca digital alimente investigaciones futuras por parte de diversos grupos de usuarios: conocedores del género caballeresco, especialistas de otros sectores de estudio y también lectores en general, interesados al contexto de producción y difusión de las obras.²⁸

Con respecto a la sostenibilidad, por otra parte, adoptamos varias estrategias. Gracias a las financiaciones obtenidas por el Proyecto de Excelencia veronés “Le Digital Humanities applicate alle lingue e letterature straniere” (2018-2022), contamos actualmente con el apoyo de los técnicos informáticos del departamento para disponer de unos servidores que nos aseguren que el portal no se convierta en un cementerio digital. Sin embargo, es evidente que muchas bibliotecas y ediciones digitales, si bien siguen siendo accesibles, necesitan una puesta al día y ajustes técnicos periódicos, debidos a los constantes cambios en

28. El tema de capturar la atención del público ha sido tratado por la comunidad científica con poca frecuencia. Peter Robinson (2005), y años más tarde Porter (2013) y Rosselli del Turco (2016), se interrogaron sobre el uso que se hace de las ediciones digitales académicas y lamentaron la falta de un amplio público con respecto a los esfuerzos humanos y económicos invertidos para su producción. Las respuestas que dieron estos investigadores trazaban un panorama de las ediciones digitales académicas que parecía todavía poco maduro para convertirse en punto de referencia para una extensa comunidad de usuarios. La nueva generación de ediciones, que se basa en una mayor interacción y difusión de los datos, quizás consentirá trazar un cuadro más riguroso y a la vez esperanzador sobre este tema. Al respecto, se vean también las reflexiones de Franzini *et al.* (2019) y Van Mierlo (2022: 117 y ss.).

los protocolos de difusión de la información, así como la permanente evolución tecnológica. Por esta razón, el flujo de trabajo que planeamos se fundamenta en la formación de jóvenes estudiantes. La transferencia de conocimiento es indispensable para contar con una nueva comunidad de usuarios que, guiados por el interés en la materia y las herramientas de las humanidades digitales, se convierten posteriormente en colaboradores del proyecto, contribuyendo a la expansión del trabajo que ya se llevó a cabo.

Desde otra perspectiva, para acrecentar la participación externa pensamos también en el reconocimiento del trabajo editorial por medio de la atribución de códigos ISBN y DOI a los recursos de la BDPM. Se trata de un paso fundamental para favorecer la incorporación al proyecto de colaboradores y estudiantes porque a los colaboradores se les asegura que los esfuerzos invertidos podrán después valorarse por parte del sistema académico. Hay que señalar que la acreditación de la actividad editorial en el medio digital ha entrado muy recientemente en los debates de la comunidad científica (Bleier 2021). Sin embargo, el reconocimiento del trabajo no parece ser todavía una preocupación de las principales instituciones europeas comprometidas en la distribución y preservación de productos editoriales. Un ejemplo es que dentro del portal ISBN no se encuentra una solución unívoca para asignarle el código a cada edición digital académica, que de momento debe considerarse una extensión de un producto editorial ya conocido como es el libro electrónico. Parece que algo se está moviendo al respecto, pero se trata de un proceso de mediación que necesita todavía cierto tiempo para asentarse.

Un último aspecto de interés, que nos ocupará en los próximos meses, es la cuestión de la accesibilidad. En concreto, entendemos la accesibilidad como un concepto extendido y de largo alcance, que abarca cuestiones como la usabilidad, la marginalidad y la inclusión (Martinez *et al.* 2019: 42 y ss.). En relación con tales asuntos, estamos definiendo un plan de trabajo que convertirá la BDPM en un espacio virtual en que se sientan representados e incluidos grupos que, de otra manera, quedarían marginados, por ejemplo adaptando el diseño de la interfaz y de productos como los eBooks a las prerrogativas de los discapacitados visuales.

Gracias a las financiaciones previstas por el nuevo proyecto de excelencia, centrado principalmente en los temas de la accesibilidad y adaptación de los recursos digitales a los principios FAIR, estamos convencidos de que se podrán alcanzar todos estos objetivos, hasta convertir la Biblioteca Digital en un lugar de conocimiento versátil y atractivo para todo tipo de usuario.

Bibliografía

- ALLÉS-TORRENT, Susanna, «“Tiempos hay de acometer y tiempos de retirar”: literatura áurea y edición digital». *Studia aurea: revista de literatura española y teoría literaria del Renacimiento y Siglo de Oro*, vol. 11 (2017), pp. 13-30, <<https://raco.cat/index.php/StudiaAurea/article/view/332408>>.
- , “Crítica textual y edición digital o ¿dónde está la crítica en las ediciones digitales?”. *Studia aurea: revista de literatura española y teoría literaria del Renacimiento y Siglo de Oro*, vol. 14 (2020), pp. 63-98, <<https://doi.org/10.5565/rev/studiaaurea.395>>.
- ALVITE-DÍEZ, María Luisa y Antonio ROJAS-CASTRO, “Ediciones digitales académicas: concepto, estándares de calidad y software de publicación”, *Profesional de la información*, v. 31, n. 2 (2021), <<https://doi.org/10.3145/epi.2022.mar.16>>.
- BAILLOT, Anne y Anna BUSCH, “Editing for Man and Machine. Digital Scholarly Editions and their Users”, *Variants*, 13 (2015), «Users of Scholarly Editions: Editorial Anticipations of Reading, Studying and Consulting», pp. 175-187.
- BARANDA LETURIO, Consolación y Eduardo RODRÍGUEZ, “Red ARACNE: retos y objetivos de un proyecto de coordinación en letras hispánicas digitales”, *Janus*, Anexo 1: «Humanidades Digitales: desafíos, logros y perspectivas de futuro», (2014), pp. 101-109.
- BAZZACO, Stefano, “El Progetto Mambrino y las tecnologías OCR: estado de la cuestión”, *Historias Fingidas*, 6 (2018), pp. 257-272, <<http://dx.doi.org/10.13136/2284-2667/89>>.
- , “Experimentos de estilometría en el ámbito de los libros de caballerías. El caso de atribución de un original italiano: Il terzo libro di Palmerino d’Inghilterra (Portonari, 1559)”, in “Prenga xascú ço qui millor li és de mon dit”. Creació, recepció i representació de la literatura medieval., (coord.) Meritxell Simó, (eds.) G. Avenzoa, A. Contreras, G. Sabaté, L. Soriano, Cilen-gua : San Millán de la Cogolla, 2021, pp. 149-166.
- BELLOMI, Paola, BOGNOLO Anna COLOMBINI; Federica, Stefano NERI, *Ciclo italiano di “Amadis di Gaula”. Collezione della Biblioteca Civica di Verona*, direzione scientifica dell’edizione fotodigitale di A. Bognolo, studi preliminari di A. Bognolo e P. Bellomi, Prefazione di A. Contò, Verona, QuiEdit, 2011-2012.
- BELLOMI, Paola, BOGNOLO Anna, Stefano NERI, Federica ZOPPI, *Repertorio delle continuazioni italiane ai romanzi cavallereschi spagnoli. Vol. 2. Ciclo di Palmerino di Oliva* (Venezia 1544-1566), Roma, Bulzoni, (en prensa)
- BLEIER, Roman, “How to cite this digital edition?”, *DHQ: Digital Humanities Quarterly*, vol. 15 (2021), Issue 3.
- BOGNOLO Anna y Stefano NERI, «Progetto Mambrino. Resultados y perspectivas», *JANUS. Estudios sobre el Siglo de Oro*, 3 (2014), pp. 68-72. <<http://www.janusdigital.es/estaticas/ver.htm?id=1>>

- BOGNOLO, Anna y Stefano BAZZACO, “Tra Spagna e Italia: per un’edizione digitale del Progetto Mambrino”, *eHumanista/IVITRA*, 16 (2019), pp. 20-36.
- BOGNOLO, Anna, «Vida y obra de Mambrino Roseo da Fabriano, autor de libros de caballerías», *eHumanista*, vol.16, 2010, pp. 77-98. (<http://www.ehumanista.ucsb.edu/>)
- , “Reti di libri, libri in rete. Libros de caballerías tra Italia e Spagna” en Maria Rosso, Felice Gambin, Giuliana Calabrese, Simone Cattaneo (eds.) *Trayectorias literarias hispánicas: redes, irradiaciones y confluencias*, Roma, AISPI Edizioni, 2018, 91-106.
- , “Attualità dello studio dei libri di cavalleria: il progetto PRIN ‘Mapping Chivalry’ nel XXI secolo”, en «Caballeresca y reescrituras (siglos XIX-XXI)», Elisabetta Sarmati y Amy Bernardi (coords.), *Orillas*, 12 (2023), pp. 485-495.
- BROOKS, Peter. *Trame: intenzionalità e progetto nel discorso narrativo*, Torino, Einaudi, 2004.
- CABO ASEGUINOLAZA, Fernando “L’Italia o l’invenzione della letteratura spagnola. Frenching Amadis. Stampa e orizzonte europeo. Ancora traduzioni. I viaggi di Diana. Picari pellegrini”, trad. Anna Bognolo, *Historias Fingidas*, 1 (2013), pp. 177-199.
- CIULA, Arianna y Cristina MARRAS, “Modelli, metamodelli e modellizzazione nelle Digital Humanities”, en Fabio Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi*, Roma, Carocci, 2023, pp. 51-65.
- COLLANTES SÁNCHEZ, Carlos M., “Red Aracne Nodus: consolidación y nuevos horizontes en las Humanidades Digitales”, en Antonio Sánchez Jiménez, Cipriano López Lorenzo, Adrián J. Sáez, José Antonio Salas (eds.), *La actualidad de los estudios de Siglo de Oro*, Kassel, Reichenberger, 2023, pp. 489-501, <http://www.reichenberger.de/Media/hd1/hd1_051.pdf>.
- DAZA, Miguel, *Corónica de don Mexiano de la Esperança, Caballero de la Fe*, ed. Ana Martínez Muñoz, Alcalá de Henares, Universidad de Alcalá / Instituto Universitario de Investigación en Estudios Medievales y del Siglo de Oro Miguel de Cervantes, 2019
- DEYERMOND, Alan “The lost Genre of Medieval Spanish Literature”, *Hispanic Review* 43, 231-59, 1975.
- FRANZINI, Greta, Simon MAHONEY y Melissa TERRAS, (2019), “Digital editions of text : Surveying user requirements in the Digital Humanities”, *ACM Journal on Computing and Cultural Heritage (JOCCH)* vol. 12, no. 1 (2019), <<https://doi.org/10.1145/3230671>>.
- LORETELLI, Rosamaria, *L’invenzione del romanzo. Dall’oralità alla lettura silenziosa*, Roma Laterza, 2010.
- LUCÍA MEGÍAS, José Manuel, “Elogio del texto digital 2.0 (¿el triunfo de la segunda textualidad?)”, en Déborah González y Helena Bermúdez Sabel (eds.), *Humanidades Digitales: Miradas hacia la Edad Media*, De Gruyter, 2019, pp. 90-110.

- , “El editor de texto ante el reto digital: elogio de la edición 2.0”, *Revista de Humanidades Digitales*, 4 (2019), pp. 93-114, <<https://doi.org/10.5944/rhd.vol.4.2019.25188>>.
- [LUJÁN, Pedro de], Leandro el Bel, ed. Stefano Bazzaco, Universidad de Alcalá, Alcalá de Henares, 2020.
- MANCINELLI, Tiziana y Elena PIERAZZO, *Che cos'è un'edizione scientifica digitale*, Carocci, Roma 2020.
- MANCINELLI, Tiziana, “Per l'edizione scientifica digitale dei ‘Documenti d'Amore’ di Francesco da Barberino: modelli, metodi e strumenti”, en Sara Bischetti y Antonio Montefusco (eds.), *Francesco da Barberino al crocevia*, De Gruyter, Berlín, 2021, p. 65-90.
- MARÍN PINA, M^a Carmen, *Páginas de sueños. Estudios sobre los libros de caballerías castellanos*, Zaragoza, Institución “Fernando el Católico”, 2011.
- MARTÍNEZ, Merisa, Wout DILLEN, Elli BLEEKER, Anna-Maria SICHANI y Aodhán KELLY, “Refining our conceptions of ‘access’ in digital scholarly editing: Reflections on a qualitative survey on inclusive design and dissemination”, *Variants*, 14 (2019), pp. 41-74, <<https://doi.org/10.4000/variants.1070>>.
- MAZZONI, Guido, *Teoria del romanzo*, Bologna, Il Mulino, 2011 [2003].
- MÜHLBERGER, Günter *et al.*, «Transforming scholarship in the archives through Handwritten Text Recognition. Transkribus as a case study», *Journal of Documentation - Emerald Publishing*, 75/5 (2019), pp. 954-976.
- NERI, Stefano. *L'eroe alla prova: architetture meravigliose nel romanzo cavalleresco spagnolo del Cinquecento*. Pisa ETS, 2007
- NERI, Stefano. *Antología de las arquitecturas maravillosas en los libros de caballerías* Alcalá de Henares Centro de estudios Cervantinos, 2007
- PAVEL, Thomas, *Representar la existencia: el pensamiento de la novela*. Barcelona Crítica, 2005.
- PENA SUEIRO, Nieves y Ángeles SAAVEDRA PLACES, «ARACNE. Red de Humanidades Digitales y Letras Hispánicas», *Historias Fingidas*, 7 (2019), pp. 407-412, <<http://dx.doi.org/10.13136/2284-2667/149>>.
- PIERAZZO, Elena y Roberto ROSSELLI DEL TURCO, “5. Critica testuale e nuovi metodi: l'edizione scientifica digitale”, en Fabio Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi*, Roma, Carocci, 2023, pp. 114-136.
- PIERAZZO, Elena, *Digital scholarly editing: Theories, models and methods*, Aldershot, Ashgate, 2015.
- , “What future for digital scholarly editions? From Haute Couture to Prêt-à-Porter”, *International Journal of Digital Humanities*, vol. 1 (2019), pp. 209-220.
- PORTER, Dot, “Medievalists and the Scholarly Digital Edition”, *Scholarly Editing*, 34 (2013), <<http://www.scholarlyediting.org/2013/essays/essay.porter.html>>.
- ROBINSON, Peter, “Where We Are with Electronic Scholarly Editions, and Where

- We Want to Be”, *Jahrbuch für Computerphilologie*, 5 (2003), 125-146, <<http://computerphilologie.uni-muenchen.de/jg03/robinson.html>>.
- , “Current Issues in Making Digital Editions of Medieval Texts —Or, Do Electronic Scholarly Editions Have a Future?”, *Digital Medievalist*, 1.1 (2005), <<http://www.digitalmedievalist.org/journal/1.1/robinson>>.
- ROJAS CASTRO, Antonio, “La edición crítica digital y la codificación TEI. Preliminares para una nueva edición de las *Soledades* de Luis de Góngora”, *Revista de Humanidades Digitales* n.1 (2016), pp. 4-19.
- RONCAGLIA, Gino, *La quarta rivoluzione. Sei lezioni sul futuro del libro*, Roma, Laterza, 2009.
- ROSSELLI DEL TURCO, Roberto, “The Battle We Forgot to Fight: Should We Make a Case for Digital Editions?”, en Matthew Driscoll y Elena Pierazzo (eds.), *Digital Scholarly Editing. Theory, Practice and Future Perspectives*, Cambridge, Open Book Publishers, 2016, pp. 219-261.
- SAHLE, Patrick, «What is a scholarly digital edition (SDE)?», en Matthew Driscoll y Elena Pierazzo (eds.), *Digital Scholarly Editing. Theory, Practice and Future Perspectives*, Cambridge, Open Book Publishers, 2016, pp. 19-39.
- SANTOS MUÑOZ, Ricardo, Eva María MOLLEJA LÓPEZ y Marta VIZCAÍNO RUIZ, “Integración de una base de datos de procedencias en el entorno MARC y de datos enlazados: el caso de la Biblioteca Nacional de España”, *RUIDE-RAe: Revista de Unidades de Información*, n. 20 (2022), III, pp. 1-11.
- SHILLINGSBURG, Peter, “How Literary Works Exist: Convenient Scholarly Editions”, *Digital Humanities Quarterly*, vol. III, n. 3 (2009), <<http://digital-humanities.org/dhq/vol/3/3/000054/000054.html>>.
- SILVA, Feliciano de, *Florisel de Niquea: partes I y II*, ed. Linda Pellegrino, Alcalá de Henares, Universidad de Alcalá, servicio de publicaciones, 2015.
- SPADINI, Elena, Francesca TOMASI y Georg VOGELER (eds.), *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, Norderstedt, BoD, 2021, <<https://kups.ub.uni-koeln.de/55227/1/Vogeler.pdf>>.
- SPENCE, Paul, “Edición académica en la era digital: modelos, difusión y proceso de investigación”, *Anuario Lope de Vega. Texto, literatura, cultura*, 20 (2014), pp. 47-83.
- STINNE GREVE RASMUSSEN, Krista, “Reading or using a digital edition? Reader roles in scholarly editions”, en Matthew Driscoll y Elena Pierazzo (eds.), *Digital Scholarly Editing. Theory, Practice and Future Perspectives*, Cambridge, Open Book Publishers, 2016, pp. 119-135.
- TROVATO, Paolo, *Con ogni diligenza corretto: la stampa e le revisioni editoriali dei testi letterari italiani, 1470-1570*, Bologna, Il Mulino, 1991.
- VAN MIERLO, Wim, “The Scholarly Edition as Digital Experience: Reading, Editing, Curating”, *Textual Cultures*, vol. 15, n. 1, Special Issue: «Provocations Toward Creative-Critical Editing» (Spring 2022), pp. 117-125, <<https://www.jstor.org/stable/48687518>>.

WATT, Ian, *Le origini del romanzo borghese: studi su Defoe, Richardson e Fielding*, Milano, Bompiani, 2017.

ZOPPI, Federica, «Reflexiones sobre la creación de una base de datos de motivos caballerescos: un desafío científico y digital», *Historias Fingidas*, Número Especial 1: «Humanidades Digitales y estudios literarios hispánicos» (2022), pp. 245-269, <<https://doi.org/10.13136/2284-2667/1103>>.

Editar en TEI teatro clásico español e inglés en los proyectos ARTELOPE y EMOTHE

Jesús Tronch Pérez

Universitat de València
tronch@uv.es

Resumen

En este trabajo se describe el tratamiento de textos electrónicos en tanto que datos en dos proyectos de bibliotecas digitales de obras del teatro clásico español e inglés que utilizan TEI en el marcado semántico de los textos: los proyectos ARTELOPE (Lope de Vega), y EMOTHE (teatro europeo 1500-1700). En esta descripción se abordan sus procesos de captura, flujo de trabajo, estructuración y publicación. Finalmente, se explica cómo editar en TEI en estos proyectos conlleva, además de la codificación semántica, juicios interpretativos y decisiones editoriales similares a los que realiza un editor que fija el texto de la obra.

Palabras clave

Edición de marcado; codificación; Text Encoding Initiative, TEI; ARTELOPE; Lope de Vega; EMOTHE; teatro europeo 1500-1700.

Abstract

Editing Classical Spanish Theater in TEI in ARTELOPE and EMOTHE.

This work describes the treatment of electronic texts as data in two projects of digital libraries focusing on classical Spanish and English theater works that use TEI for semantic markup of the texts: the ARTELOPE projects (Lope de Vega) and EMOTHE (European theater 1500-1700). This description covers their capture processes, workflow, structuring, and publication. Finally, it explains how editing in TEI in these projects involves, in addition to semantic encoding, interpretative judgments, and editorial decisions similar to those made by an editor who establishes the text.

Keywords

Markup editing; encoding; Text Encoding Initiative, TEI; ARTELOPE; Lope de Vega; EMOTHE; early modern European theatre.

En su reciente libro *Theater as Data*, Miguel Escobar (2021: 75-93) analiza varias facetas en las que el estudio del teatro se enriquece con la ciencia de los datos (en el sentido de información almacenada y procesada electrónicamente), y en la clasificación de estas facetas, destina un capítulo a las “palabras como datos” junto a las de “relaciones”, “movimiento” y “lugar”.¹ En estas páginas describiré un tratamiento de las palabras en tanto que datos en el teatro clásico español e inglés, en concreto cuando estas palabras se “editan en TEI” en dos proyectos que se ocupan de la difusión de obras teatrales de la época a través de textos electrónicos, y en los que participo en varios órdenes y niveles como editor, codificador, revisor y coordinador.

Aún a riesgo de ser innecesario, recordemos que “TEI” es el acrónimo del consorcio Text Encoding Initiative, que elabora y actualiza unas *Guidelines* o directrices para la codificación semántica, en lenguaje de marcado XML, de textos humanísticos electrónicos (TEI Consortium 2023). Todo texto electrónico, en tanto que conjunto de datos que procesará un ordenador, necesita un modelo de datos que describa cómo se organiza y se representa la información textual que transmite. Las directrices TEI son un modelo de datos (Escobar 2021: 78) que establecen un esquema de reglas de codificación que identifica y describe semánticamente elementos textuales, sus categorías y sus interrelaciones, y un vocabulario controlado para los códigos o marcas insertados en el texto digital que delimitan esos elementos.

Los dos proyectos en cuestión son ARTELOPE, sobre las comedias de Lope de Vega, y EMOTHE, sobre el teatro europeo de los siglos XVI y XVII (en concreto las tradiciones italiana, francesa, inglesa, española y portuguesa).² Además de ofrecer bases de datos en acceso abierto con información sobre las obras teatrales, ambos proyectos producen ediciones digitales de las obras en sendas bibliotecas en línea y datos estadísticos sobre intervenciones de los personajes, estructura de la obra, y en su caso, estructuras métricas.

Una cuestión terminológica previa

En las páginas que siguen, para referirme a los textos electrónicos codificados producidos por los proyectos, utilizo el término “edición digital”, dado que, siguiendo la definición de Sahle (2016: 26-28), no son una mera digitalización del texto de una edición previa (generalmente impresa) sino que ofrecen

1. Esta publicación es parte del proyecto I+D+i «Teatro español y europeo de los siglos XVI y XVII: patrimonio y bases de datos», referencia PID2019-104045GB-C54 (acrónimo EMOTHE), financiado por MICIN/AEI/10.13039/501100011033.

2. ARTELOPE, dir. Joan Oleza, Universitat de València, <https://artelope.uv.es/>; EMOTHE, dir. Jesús Tronch y Joan Oleza, Universitat de València, <https://emothe.uv.es/>.

funcionalidades propias del medio digital derivadas principalmente del marcado semántico del texto. Y utilizo el término “editor digital” para referirme a la persona o personas responsables de esa edición digital (edición y codificación del texto electrónico para ARTELOPE y para EMOTHE), sin presuponer que este editor es la misma persona que fija el texto de la obra (en papel o en soporte electrónico). Como señalo más adelante, la mayoría de las ediciones digitales en estos proyectos son una representación digital de ediciones previas, y en estos casos pueden denominarse “metaediciones”, siguiendo a Portela (2022: 101).

Empleo deliberadamente los términos derivados de “editar”, teniendo en cuenta las acepciones “publicar”, “adaptar un texto” y “modificar un documento o archivo”, como recoge el Diccionario de la Real Academia, a las que añado “representar”. Las propias directrices TEI se definen como recomendaciones de maneras adecuadas de *representar* [énfasis mío] aquellas características de recursos textuales que necesitan que se identifiquen explícitamente para facilitar su procesamiento por programas informáticos (“About”). Como explica Burnard (2014: par. 3), el lenguaje de marcado XML para textos electrónicos proporciona una manera simple de *representar* [énfasis mío] datos estructurados como una secuencia lineal de datos de caracteres y de etiquetar partes concretas de esa secuencia con etiquetas con nombre, para indicar una semántica o una función estructural de las mismas.

Esta “representación” en TEI suele denominarse marcado o codificación (*markup* o *encoding* en inglés), y *encoder* o codificador a quien se encarga de ella. Pero prefiero aprovechar la polisemia del término “editar”, porque en ARTELOPE y en EMOTHE el tratamiento del texto para representarlo en un archivo XML-TEI puede conllevar intervenciones en el texto digital que se suman a la codificación en tanto que simple anotación o descripción semántica de elementos y características textuales que interesan a los proyectos, y, como explicaré en el apartado “Editar en TEI”, a menudo conlleva juicios interpretativos y decisiones editoriales similares a los que realiza quien establece el texto.

Para justificar esta elección léxica de “edición”, “editor”, “editar”, traigo a colación ejemplos provenientes de proyectos editoriales análogos. En la serie de ediciones electrónicas shakespearianas *Folger Digital Texts*, ahora renombrada como *The Folger Shakespeare*,³ los encargados de la transformación del texto impreso a distintos formatos electrónicos, Michael Poston y Rebecca Niles, definen su responsabilidad como “Edited for XML and encoded by”, para distinguirla de la de los editores críticos del texto que se publicó primeramente en papel, Paul Werstine y Barbara Mowat. Esta frase (que reúne “editar” y “codificar”) es la que consta en los metadatos de los archivos XML de

3. *The Folger Shakespeare*, Folger Shakespeare Library, <<https://www.folger.edu/explore/shakespeares-works/>>;

cada obra.⁴ Además, en la actual página web de *The Folger Shakespeare* que informa sobre los diferentes formatos en los que los usuarios se pueden descargar los textos (“Usage”), se describe a Michael Poston como “Digital Editor” (junto al de “Encoding Architect”), mientras que en la versión anterior se describía como “Digital Texts Editor and Encoding architect” y a Rebecca Niles como “Digital Texts Editor and Interface architect” (“About”). Es interesante observar que al final de su ensayo sobre *Folger Digital Texts*, Niles y Poston (2016: 141) emplean el término compuesto “editor-encoders” después de haber recurrido en varias ocasiones a la pareja “editors and encoders” y “encoders and editors” (2016: 118-119). Otro uso de “editor” es el que observamos en la frase “markup editor” en el proyecto *Digital Victorian Periodical Poetry*,⁵ en cuya documentación se define “markup editor”, como el que realiza la codificación del texto o los metadatos en un determinado formato o lenguaje de marcado SGML, HTML, o XML (Chapman 2023). Por último, cuando el estándar usado en la codificación es el propuesto por el consorcio Text Encoding Initiative, no es infrecuente encontrar el término “TEI edition”, “edition in TEI”, o similares.⁶

Descripción de los proyectos ARTELOPE y EMOTHE

El Proyecto ARTELOPE ofrece en acceso abierto una base de datos que permite explorar diversos aspectos de 433 comedias de, o relacionadas con, Lope de Vega (no solo las de autoría fiable sino también las probables y dudosas, y algunas atribuidas que son de especial interés aunque no le pertenezcan), y está creando una biblioteca en línea de ediciones digitales de estas obras. La Biblioteca Digital que acompaña la base de datos, que inició su andadura una vez esta última estuvo plenamente operativa (Oleza 2014: 219-220), tiene como objetivo que los usuarios puedan acceder fácilmente, sin salir de la interfaz, a un texto electrónico con grafía y puntuación modernizada, a estadísticas sobre aspectos estructurales de la obra (versos y fragmentos en prosa en los distintos actos o jornadas, número de acotaciones y de apartes), sobre porcentajes de interven-

4. En concreto en la declaración de responsabilidad (etiquetada con <respStmt>) en la sección destinada a la descripción del archivo (<fileDesc>), dentro del encabezamiento o <teiHeader> del propio archivo electrónico.

5. *Digital Victorian Periodical Poetry Project*, dir. Alison Chapman, Edition 0.98.8beta, University of Victoria, 30/06/2023, <https://dvpp.uvic.ca/index.html>.

6. Véase *Helsinki Corpus TEI XML Edition* (2011), Fukushima y Bourrier (2019-2020), Fukushima, Bourrier y Parker (2022: abstract, par. 3 y 10), Beshero-Bondar (2018), y los proyectos *Digital Dinah Craik* (University of Calgary, <https://www.digitaldinahcraikproject.org/>), *Mabbott Poe* (University of Iowa, <https://mabbottpoe.org/>), *The Peter Still Digital Edition* (Rutgers University, <https://stillpapers.org/about/>), y *XML Helsinki Corpus Browser* (University of Helsinki, <https://helsinki Corpus.arts.gla.ac.uk>).

ción de los personajes, y sobre la versificación; y además que puedan realizar búsquedas de palabras y frases en todo el corpus que reúne la base de datos. Hasta la fecha, la Biblioteca Digital ARTELOPE ofrece en acceso abierto ediciones digitales de 334 títulos.

El Proyecto EMOTHE, y en concreto su Biblioteca Digital, publica en acceso abierto ediciones modernizadas de una selección de obras teatrales de las tradiciones italiana, francesa, española, portuguesa e inglesa de los siglos XVI y XVII, así como traducciones y adaptaciones (históricas o contemporáneas) y los datos estadísticos de todos los textos. Como indica Oleza (2013: 152), este proyecto representa una expansión en dirección europea de la experiencia del grupo de investigación ARTELOPE, en cuyo seno surgió EMOTHE. En esta Biblioteca Digital y en su base de datos, el teatro español está principalmente representado por las obras de la colección CANON 60 que elaboró el grupo TC/12 (Oleza 2015), pero también por otras obras de interés desde una perspectiva europea (por ejemplo, *Don Lope de Cardona*, que tiene adaptaciones interlingüísticas en *Don Lope de Cardone* de Jean Rotrou y en *The Young Admiral* de James Shirley), o porque han sido fruto de un trabajo Final de Máster o de tesis doctorales (por ejemplo, *La estrella de Sevilla*, editada por Nàdia Revenga). Hasta la fecha, la Biblioteca ofrece unas 350 ediciones entre originales de las cinco tradiciones (180) y traducciones y adaptaciones (160). En el caso del teatro español, unas 68 originales y 42 traducciones.

Las bibliotecas digitales de ARTELOPE y de EMOTHE son “proyectos de biblioteca y archivo”, una de las siete categorías de proyectos de humanidades digitales que distingue Leonhardt (2014: 6-7), al igual que pueden clasificarse proyectos como LEMDO (acrónimo de Linked Early Modern Drama Online, que reúne proyectos y antologías de teatro anglófono de la época y que se desarrolla en la University of Victoria, en Canadá),⁷ *Folger Shakespeare* (antes *Folger Digital Texts*, con ediciones críticas modernizadas de la obra de Shakespeare) o *EarlyPrint* (anteriormente *Shakespeare His Contemporaries* y ahora abarcando textos no teatrales).⁸ ARTELOPE y EMOTHE llevan a cabo una codificación de los textos dramáticos en lenguaje de marcado XML según las directrices TEI, que es una práctica bastante generalizada en los proyectos centrados en textos humanísticos.

Los dos son proyectos colaborativos (como LEMDO), en los que intervienen muchos y distintos “editores digitales” de los textos a través de plataformas en las que se puede trabajar remotamente mediante programas que centralizan la información en servidores de las respectivas universidades. Son proyectos concebidos para tratar gran número de textos de obras de teatro de los siglos XVI y

7. LEMDO, dir. Janelle Jenstad, University of Victoria, <https://lemdo.uvic.ca/>.

8. *EarlyPrint*, Northwestern University and Washington University in St. Louis, <https://earlyprint.org/>.

xvii y los metadatos sobre las obras. Y finalmente, son proyectos cuyos textos electrónicos están en acceso abierto.

Seguiré la descripción en torno a los tres aspectos del proceso ecdótico que distingue Allés-Torrent (2020: 75-78): captura, estructuración y publicación, a los que añadiré un subapartado sobre el flujo de trabajo y software.

Captura y procedencia de los textos

Las bibliotecas digitales de ARTELOPE y EMOTHE están concebidas como una plataforma para facilitar el conocimiento de las obras teatrales a través de textos electrónicos y su codificación semántica en XML-TEI, al servicio de otros usuarios e investigadores. Para este fin, la gran mayoría de las ediciones digitales capturan el texto establecido en una edición fuente externa, mientras que un porcentaje menor son ediciones en las que el texto está establecido por investigadores o colaboradores propios de los proyectos. Incluso en bastantes de estos casos, los editores del texto (los que fijan el texto, críticamente o no), no son los “editores digitales”. Generalmente esa captura de textos externos a los proyectos se realiza mediante una transcripción a soporte informático encargada, o, en el caso de ARTELOPE, se realiza en el proceso de modernización de la grafía llevado a cabo por el propio editor digital responsable de la obra en cuestión.

La procedencia de los textos es diversa. En ARTELOPE, algo más de un 60% proviene de los textos editados por la Real Academia Española, por estudiosos como Emilio Cotarelo, Marcelino Menéndez Pelayo, y en menor medida Eduardo Juliá, textos que en bastantes casos son la única edición moderna disponible y que cuentan con la ventaja de estar en dominio público. Un 20% son ediciones relativamente recientes de estudiosos, críticas o no, firmadas por hispanistas como, por ejemplo, Alonso Zamora Vicente (*La dama boba*), María Grazia Profetti (*El caballero de Olmedo*), Hélène Tropé (*Los locos de Valencia*), Alejandro García Reidy (*El castigo sin venganza*), Guillermo Serés (*El villano en su rincón*), o Abraham Mañoral (*El ausente en su lugar*). En este grupo destacan por el número de ediciones empleado (un 10% del total) Jesús Gómez y Paloma Cuenca, con sus ediciones para la Fundación Castro. En los casos de ediciones con aparato crítico de variantes y enmiendas, este no se reproduce en la Biblioteca Digital. El 20% restante corresponde a textos establecidos por investigadores propios de los proyectos, ediciones que, como las de la Colección CANON 60 o las de la Fundación Castro, modernizan la grafía y la puntuación y, en una gran mayoría, no ofrecen aparato crítico o explicación de la transmisión histórica del texto y de los procedimientos editoriales.

La elección de la procedencia de la gran mayoría de las ediciones en la Biblioteca Digital ARTELOPE ha recibido juicios desfavorables por no ser ediciones críticas sino actualizaciones de ediciones anticuadas de Menéndez Pelayo, Cotarelo y Mori, etc. Respondiendo a estos juicios, el director de ARTELOPE, Joan Oleza, ha señalado que efectivamente la gran mayoría no

son ediciones críticas “porque nunca han querido serlo”, puesto que para ARTELOPE “lo fundamental [es] producir en el menor tiempo el mayor número posible de ediciones digitales, realizadas sobre fuentes solventes para facilitar los usos digitales de los textos”, como el análisis estilométrico del grupo ETSO de la Universidad de Valladolid, o el reconocimiento digital automático de textos manuscritos llevado a cabo por el centro Pattern Recognition and Human Language Technology de la Universidad Politécnica de Valencia;⁹ y para ofrecer el análisis de las estadísticas de cada obra como datos procesables: “a estos propósitos sirven de la misma manera ediciones críticas y ediciones no críticas”.¹⁰ Este imperativo temporal hace inviable la elaboración, como pauta general, de ediciones críticas propias que requiere la dedicación completa de un investigador durante, al menos, un año, y que a efectos de esta investigación digital es un trabajo improductivo.

En EMOTHE, los textos provenientes de una edición fuente externa constituyen un 75%, tanto en títulos originales en los respectivos idiomas de cada tradición como en traducciones. Los de títulos originales son: (1) ediciones críticas ya publicadas de las que se obtiene permiso, en papel (como la de *Don Lope de Cardona*, de Eugenia Fosalba), o telemática (como las procedentes de *Digital Renaissance Editions*, o *Richard Brome Online*); (2) ediciones firmadas por estudiosos sin que cuenten con aparato crítico consistente en introducción y notas de enmiendas y variantes (como *A Chaste Maid in Cheapside* del dramaturgo Thomas Middleton, en edición de Bryan Loughrey y Neil Taylor, impresa en papel, o como las telemáticas de *Théâtre Classique*); y (3) recodificación de ediciones digitales externas, como las de las colecciones CANON 60, *Folger Shakespeare*, o *Internet Shakespeare Editions*.¹¹

De las traducciones y adaptaciones de procedencia externa (un 72% del total de traducciones y adaptaciones), no solo se incluyen traducciones ya publicadas, con la autorización correspondiente de las que no están en dominio público, sino también traducciones inéditas, de las que se ha tenido noticia y se han rescatado del “cajón del olvido”, con el permiso de sus autores.

9. ETSO: *Estilometría aplicada al Teatro del Siglo de Oro*, dir. Álvaro Cuéllar y Germán Vega García-Luengos, Universidad de Valladolid, <<http://etso.es/>>; Pattern Recognition and Human Language Technology Research Center, Universitat Politècnica de València, <https://www.prhlt.upv.es/handwritten-text-recognition/>.

10. Comunicación privada de Joan Oleza incluida en informe remitido al Ministerio de Ciencia e Innovación sobre el proyecto de investigación EMOTHE, referencia PID2019-104045GB-C54.

11. Hay casos en los que el responsable de la fuente externa es a su vez el responsable de la edición digital en EMOTHE, como ocurre con José Camões y los proyectos que dirige en el Centro de Estudios de Teatro, *Teatro de Autores Portugueses do Séc. XVI*, <<http://www.cet-e-quinientos.com/>>; y *Teatro de Autores Portugueses do Séc. XVII*, <<http://www.cet-e-seiscentos.com/>> o con Alejandro García Reidy, responsable de la codificación para EMOTHE de su edición crítica de *El castigo sin venganza* publicada en papel por Milenio en 2009.

Estructuración: modelo de datos y personalización TEI

En tanto que modelo de datos, las directrices TEI prevén muchas etiquetas, con sus atributos y posibles valores, para codificar la semántica de determinados elementos del texto, de manera que puedan satisfacer un amplio espectro de ediciones, por lo que cada edición o cada proyecto necesita centrarse en una selección de los elementos a codificar, es decir, necesita hacer una *customization* o personalización. Los proyectos ARTELOPE y EMOTHE, en tanto que dedicados a textos teatrales que usan el verso, confeccionan su propio subconjunto de etiquetas TEI a partir de los módulos “6. Verse” y “7. Performance Texts”, además de los módulos comunes a todo documento TEI; y en el caso de ediciones críticas en EMOTHE, a partir del módulo “12 Critical Apparatus”.

Los elementos del texto a codificar son los personajes del drama (elenco), las divisiones de la obra principalmente en actos y escenas, y también en paratextos de diversos tipos, las acotaciones, los parlamentos, las didascalias de interlocutor, versos (incluyendo partes iniciales, medias y finales de versos compartidos), párrafos y líneas de prosa, las palabras o segmentos del diálogo que se dicen en “aparte”, los segmentos textuales que reciben énfasis lingüístico o retórico (y que se visualizan como cursiva), las estructuras métricas o estrofas (en EMOTHE solo en las obras españolas y portuguesas), notas al texto (tanto las que están presentes en el texto utilizado para la edición digital, o las aportadas por el editor digital del proyecto); y aparato crítico (solo en EMOTHE y en aquellas ediciones críticas realizadas por investigadores del proyecto o que han constituido un trabajo Final de Máster o una tesis doctoral).

La lista de los personajes se codifica como un paratexto (y, por tanto, dentro de la sección frontal del cuerpo del texto marcada con el elemento <front>), con el elemento <castItem> en una lista de personajes <castList>, con los consiguientes elementos <role> para el nombre del personaje (por ejemplo “Feliciano”) y <roleDesc> para la descripción de este papel (por ejemplo, “caballero”). Otros proyectos, como *LEMDO*, codifican las *dramatis personae* en la descripción del perfil no bibliográfico del texto (elemento <profileDesc>) dentro de la cabecera con los metadatos del archivo (el <teiHeader>), y en concreto con la etiqueta <listPerson>, dentro de la descripción de los participantes <particDesc>. Las discrepancias que puede haber entre el nombre del personaje que aparece en la lista de *dramatis personae* y el nombre como didascalia de interlocutor en cada parlamento se salvan al asociar cada uno de estos a un identificador XML único que el responsable de la edición digital asigna a cada personaje.

La codificación en TEI mediante un identificador único de los interlocutores permite que un personaje que varía su nombre en la didascalia de interlocutor a lo largo de la obra tenga todos los parlamentos asociados al mismo personaje y permite distinguir a qué personajes distintos corresponde en cada momento una misma expresión del interlocutor (por ejemplo, en *Las mocedades del Cid: comedia segunda*, la didascalia “DON DIEGO” puede referirse a Don

Diego Ordóñez de Lara o a uno de los cinco hijos de Arias Gonzalo). Asimismo, se puede identificar los personajes en didascalias de interlocutores generales tipo “Todos”, “All”, “Ambos” “Both”, o “Voces”, “Dentro”, “Voices within”.

Se codifican los versos de un parlamento dentro del elemento <lg> (grupo de versos) en todos los casos, incluido cuando no está previsto identificar estas agrupaciones con determinadas estrofas (por ejemplo, *blank verse* en teatro inglés).

Un elemento que ARTELOPE y EMOTHE codifican de manera singular (al menos desconozco otros proyectos que lo hagan), es la sección o segmento de un parlamento que se dice en “aparte”. Este singular aspecto se codifica como <seg type=“aside”>: con la etiqueta del elemento <seg>, dando el valor “aside” al atributo @type. Las directrices TEI no contemplan una etiqueta específica para “aside”, ni tampoco recogen “aside” como valor que defina un tipo de acotación junto a las categorías “delivery”, “business”, “entrance”, etc.¹²

La base de datos en FileMaker de ARTELOPE no está programada para la codificación del aparato crítico según las directrices TEI, mientras que la de EMOTHE, desarrollada con posterioridad, sí que facilita esta opción. El método TEI para enlazar texto y aparato crítico es el de *parallel segmentation* (en contraste con, por ejemplo, LEMDO, que utiliza el método *double end-point attachment* con un archivo externo que contiene el aparato crítico).

Las notas, marcadas con el elemento <note> dentro del propio archivo (no en un archivo externo), se pueden asociar a cualquier componente del texto (personajes del elenco, interlocutores, acotaciones, segmentos en el diálogo), y clasificar según el valor seleccionado para el atributo @type. Los tipos de nota son: nota del autor, del traductor, del editor crítico o una nota propia como editor digital.

ARTELOPE y EMOTHE no codifican elementos estructurales inferiores a la línea de verso o de prosa, o a la acotación. Tampoco se agrupan segmentos de diálogo dispersos en varias unidades estructurales (por ejemplo, texto de canciones repartido entre varios parlamentos), ni personajes del elenco que estén relacionados; los encabezamientos del tipo “Canción”, “Carta”, están tratados como acotación en vez de estar etiquetados con el elemento <head>. Entre los signos de puntuación, las comillas (dobles o simples) y los corchetes cuadrados no se codifican según su semántica o función, y lo que se visualiza como cursiva, se codifica sistemáticamente con <emph>.

Una selección de obras reciben una edición “avanzada” en tanto que se codifican, con vistas a ulteriores estadísticas, aspectos como antropónimos, topónimos (aludidos o relacionados con la acción de la obra), referencias mitológicas, oficios, citas literarias, latinismos, referencias a lugares del escenario,

12. Una opción habría sido añadir un nuevo atributo, de manera análoga a como se hizo con el atributo “asonancia” en el proyecto ReMetCa de métrica medieval castellana (González-Blanco y Rodríguez 2015, para. 11).

elementos escenográficos, atrezo, vestuario, gestos, movimiento, y efectos especiales. El elemento TEI empleado para estos aspectos del texto es el *referencing string* <rs>, con los correspondientes valores en el atributo @type para indicar los diferentes aspectos.

Para las referencias al texto por numeración, ARTELOPE numera todos los versos (no los textos en prosa, como es el caso de cartas) de manera continua sin reiniciar el cómputo en cada acto o jornada. En EMOTHE, dado que se trata de una biblioteca que recoge ediciones de distintas tradiciones teatrales, el editor digital tiene varias opciones: con respecto a la segmentación de la numeración, por obra, actos/jornadas, y escenas; y con respecto a los elementos numerados, versos solo, versos y prosa, e intervenciones o parlamentos. Los parlamentos en prosa se pueden codificar de distinta manera, como un solo párrafo, o permitiendo que cada línea tenga un número (según la numeración del texto fuente o la decidida por el editor textual).

A partir de los textos editados en TEI, se obtienen datos estadísticos sobre (1) intervenciones de los personajes (cuántas intervenciones están asignadas a cada personaje que habla en la obra, y en su caso cuántos párrafos, versos o versos compartidos), (2) estructura de la obra (número de secciones en que se divide la obra, distribución del número de versos o elementos de división del diálogo en cada sección, número de acotaciones, número de apartes), y (3) en su caso, estructura métrica (desglose de formas estróficas, con el cómputo de versos y de estrofas para cada forma; y listado secuencial de las estrofas, indicando el rango de versos).

En resumen, ARTELOPE y EMOTHE aplican a la gran mayoría de las ediciones digitales un marcado que podemos calificar como básico pero suficiente para los propósitos de ambos proyectos de conseguir codificar un gran número de textos diversos, a cargo de diferentes editores digitales para los que, en su gran mayoría no necesitan conocer las directrices TEI, y a través de una plataforma que no requiere el uso de una herramienta de edición de XML (como oXygen). La codificación semántica de componentes del texto no es tan densa como la realizada en LEMDO, que marcan signos ortográficos e introduce elementos de anclaje para enlazar con otras partes del texto o con otros archivos (Jenstad 2023: 242-243). Pero el marcado en ARTELOPE, EMOTHE y LEMDO no es tan complejo como el que realiza *Folger Digital Texts*, ahora *Folger Shakespeare*, en el que las unidades mínimas de codificación son la palabra, el signo de puntuación y el espacio en blanco; o el de *EarlyPrint*, (anteriormente *Shakespeare His Contemporaries*), cuyo marcado denso contempla la identificación de la categoría gramatical de cada palabra.¹³ La personalización TEI en ARTELOPE y EMOTHE

13. *The Folger Shakespeare*, Folger Shakespeare Library, <https://www.folger.edu/explore/shakespeares-works/>; *EarlyPrint*, Northwestern University and Washington University in St. Louis, <https://earlyprint.org/>.

es más parecida a la de *Early English Books – Text Creation Partnership*,¹⁴ o la de la colección CANON 60 realizada por la Biblioteca Virtual Cervantes, si bien a diferencia de estas la codificación de ARTELOPE y EMOTHE enriquece los textos con la asignación de cada parlamento al interlocutor correspondiente, asociación necesaria para generar los datos estadísticos.

Flujo de trabajo y software

El flujo de trabajo y el *software* empleado en ARTELOPE y EMOTHE es muy similar, ya que este último surgió en el seno del primero, como he indicado anteriormente. En ambos proyectos el responsable de la edición digital emplea en primer lugar un editor de texto plano o un procesador de textos, y completa la codificación en el programa de gestión de la base de datos, el *software* propietario FileMaker.¹⁵ Una vez acabada la codificación, FileMaker genera un archivo XML-TEI, que es de nuevo validado con el programa oXygen, editor XML propietario. Para asegurar la actualización continua, los responsables de la edición digital tienen conectado el programa FileMaker constantemente con el servidor alojado en la Universitat de València (Muñoz Pons 2013: 487).

La primera fase del tratamiento del texto electrónico consiste en un marcado previo, realizado en un editor de texto plano o en un procesador de textos, con etiquetas propias, insertadas solo al principio del elemento estructural o característica del texto, sin necesidad de introducir una etiqueta de cierre, ya que FileMaker está programado para detectar en una misma línea la marca de párrafo o una marca preliminar distinta como punto de inserción de la etiqueta de cierre. Sí que se debe marcar el inicio y final del segmento de diálogo que se dice en “aparte” y las palabras o frases enfatizadas mediante cursiva. Este sistema de trabajo se concibió para que los numerosos y distintos responsables de la edición digital no necesitaran conocer o recibir entrenamiento en el estándar TEI ni adquirir un editor de XML como oXygen.

Una vez el texto está “premarcado”, se inserta en la base de datos FileMaker, los actos y jornadas y distintos paratextos por separado. El programa distribuye los elementos “premarcados” en distintos campos o tablas de la base de datos. Se sigue con la codificación a través de una ventana de edición del FileMaker, en la que se van insertando etiquetas mediante botones, ventanas de diálogo y menús

14. *Early English Books Online – Text Creation Partnership*, The University of Michigan Library, <https://quod.lib.umich.edu/e/eebgroup/>.

15. Cabe señalar que la base de datos ARTELOPE empezó a desarrollarse en 1998, cuando programas similares de código abierto no disponían de las mismas funcionalidades. En el desarrollo técnico de la base de datos intervino Arturo Silvestre Mora, de quien tomó el relevo Carlos Muñoz Pons.

desplegables en una interfaz sencilla. En este proceso, se asignan identificadores a cada parlamento y se pueden añadir notas y distinguir los tipos de acotaciones.

Una vez finalizada la codificación por parte del editor digital, este acciona el proceso por el que FileMaker revisa la estructura y jerarquía de la codificación de cada acto o jornada; y en el caso de aparecer errores identificados por FileMaker, los corrige y vuelve a accionar la revisión de la estructura. Una última operación asigna automáticamente la numeración de líneas (en verso y/o prosa) o de parlamentos así como distintos identificadores XML a lo largo del archivo. En ARTELOPE este archivo XML-TEI contiene identificadores `xml:id` para cada verso, mientras que en EMOTHE también tienen identificadores las divisiones en actos y escenas, los parlamentos, acotaciones y párrafos en prosa.

A partir de entonces, el proceso pasa a manos de los técnicos informáticos o del responsable del proyecto para extraer el archivo XML-TEI de la base de datos y validarlo con el esquema TEI-All. Este archivo XML-TEI (que constituye una “edición TEI”) es el fichero matriz, a partir del cual se generan archivos posteriores: archivo HTML-PHP, mediante la transformación con el lenguaje de hojas de estilo extensible XSL, para la visualización del texto de la obra en una edición hipertextual disponible en la web; y, próximamente, archivo en formato EPUB; y un archivo XML con los datos estadísticos (tras un análisis realizado con una aplicación desarrollada *ad hoc*), que posteriormente se transforma con XSL en un archivo HTML asociado accesible desde la edición hipertextual.

Publicación

En el apartado de publicación, distinguimos entre la Biblioteca Digital en la web y los repositorios donde se depositan en acceso abierto los archivos XML-TEI de las obras teatrales. Estos últimos, así como los archivos de los datos estadísticos, están disponibles en la plataforma GitLab;¹⁶ mientras que en EMOTHE, se pueden descargar individualmente desde el índice de la Biblioteca Digital.¹⁷

Como se ha apuntado anteriormente, para su visualización en la web, los dos proyectos transforman los archivos matriz XML-TEI en archivos HTML con extensión `.php` porque se recurre a funcionalidades dinámicas del lenguaje de programación PHP para que, a partir de archivos HTML-PHP que combinan código de ambos lenguajes, sea el servidor el que remita archivos en HTML al navegador para visualizar su contenido.

16. La de ARTELOPE en <https://gitlab.com/artelope1/ARTELOPE>, la de EMOTHE en <https://gitlab.com/emothe/emothe-research-project>.

17. Y sin tener por el momento la totalidad de los archivos disponibles, desde los repositorios abiertos ZENODO (el servicio operado por el CERN), <https://zenodo.org/communities/emothe>, y RODERIC (el repositorio institucional de la Universitat de València), <https://roderic.uv.es/>.

En estos momentos estamos considerando la posibilidad de adoptar las recomendaciones del *Endings Project*,¹⁸ que ofrece estándares recientes para preservar los datos (Jenstad 2023: 241): estándares viables a largo plazo como HTML5, JavaScript, CSS, en un sitio de internet estático que no dependa de *software* situado en un servidor (sin bases de datos, sin PHP, sin Python, ni funcionalidades de búsqueda), para que las ediciones se puedan depositar en bibliotecas, servidores y ordenadores personales, y puedan utilizarse incluso sin conexión a internet.¹⁹

Las notas de comentario o glosas, así como las de variantes, se muestran por medio de ventanas emergentes que el usuario tiene que llamar y cerrar, como es habitual en ediciones web. Además, las ediciones HTML-PHP contienen funcionalidades que permiten al usuario optar (mediante selectores en el panel de navegación) por visualizar de manera destacada las acotaciones, los versos compartidos, los apartes y la métrica (como se muestra en la siguiente imagen).

The screenshot shows the EMOTHE digital library interface. At the top, it identifies the collection as 'COLECCIÓN EMOTHE' and the play as 'EL CASTIGO SIN VENGANZA' by 'Félix Lope de Vega y Carpio'. A sidebar on the left contains a navigation menu with options: 'Cerrar panel X', 'Índice de navegación:', 'Datos de la edición', 'Dedicatoria', 'Prólogo', 'Elenco', 'Acto 1', 'Acto 2', 'Acto 3', 'Marcas visuales:' (with checkboxes for 'Acotaciones', 'Apartes', 'Versos partidos', and 'Ocultar métrica' which is checked), and 'Estadísticas'. The main content area shows a scene titled 'Vase' with a 'FEDERICO' section. The text includes: '¿Qué buscas, imposible pensamiento? Bárbaro, ¿qué me quieres? ¿Qué me incitas? ¿Por qué la vida sin razón me quitas, donde, volando, aun no te quiere el viento? Detén el vagaroso movimiento, que la muerte de entrambos solicitas; déjame descansar y no permitas tan triste fin a tan glorioso intento. No hay pensamiento, si rindió despojos, que sin determinado fin se aumente, pues, dándole esperanzas, sufre enojos. Todo es posible a quien amando intente, y sólo tú naciste de mis ojos para ser imposible eternamente.' Verse numbers 1800, 1805, and 1810 are visible. Below this is the 'Casandra entre' section with a 'CASANDRA' section: 'Entre agravios y venganzas anda solícito amor después de tantas mudanzas, sembrando contra mi honor mal nacidas esperanzas.' Verse number 1815 is visible.

Fig. 1: Captura de pantalla de la edición de *El castigo sin venganza* en la Biblioteca Digital EMOTHE con la opción de “mostrar métrica” seleccionada.

18. *The Endings Project*, University of Victoria, <<https://endings.uvic.ca/index.html>>.

19. Se han adherido a estas directrices proyectos como *LEMDO* o *Digital Victorian Periodical Poetry Project*, que ofrecen ediciones estáticas que viajan con sus propios metadatos, pero que, en el caso de *LEMDO*, pueden integrarse en nuevas antologías (Jenstad 2023: 249)

En EMOTHE, los textos de las obras originales y las de sus traducciones se ofrecen también en edición de textos paralelos, aun cuando en la mayoría de los casos las traducciones no se realizaron exactamente a partir de la edición del original usada en EMOTHE.

Editar en TEI

Según las características del texto fuente a representar digitalmente, editar en TEI para ARTELOPE y EMOTHE puede consistir solo en el cotejo del texto transcrito y en su codificación semántica, pero también puede conllevar distintos tipos y grados de intervención en el texto, así como juicios de interpretación y decisiones editoriales similares a los realizados en una edición tradicional. Como han observado Niles y Poston (2016: 136), la práctica de codificar el texto de las ediciones shakespearianas de Mowat y Werstine es similar a la de editar estos textos shakespearianos, ya que ambas dependen del desarrollo de una comprensión abstracta del texto y de una preparación del mismo, ambas idean cómo representar mejor características significativas del texto, y ambas tienen en cuenta las diversas capas del texto: su contenido, su modelo de datos y su hoja de estilo.

Tanto si se está editando en el sentido de fijar el texto de una obra de manera digitalmente nativa, como preparando una “metaedición” digital de un texto ya establecido (propio o por otra persona), editar en TEI supone pensar en el texto más allá del modo analógico en el que lo visualizamos (en papel o en una pantalla) y considerar sus componentes estructurales y otras características de interés como datos; supone, como señala Allés-Torrent (2020: 75), ser consciente “de lo que implica trabajar con datos informáticos”, con la “materialidad digital” de la edición electrónica.

Un ejemplo ilustrativo de lo que implica editar en TEI un texto entendido como datos los podemos encontrar en los casos de canciones o cartas para los que el texto fuente impreso no indica explícitamente quién canta o quién lee la carta. En *El ausente en su lugar* de Lope de Vega, verso 1278 (en la edición crítica de Abraham Madroñal, que es la que usa ARTELOPE para la Biblioteca Digital), Carlos lee una carta después de un parlamento de Marquina:

MARQUINA

Tú verás
que Elisa no puede más,
pues eres, Carlos, discreto.

Lea Carlos

«Si una mujer principal,
Porque a su padre obedece.[...]»

Madroñal ha representado como acotación “*Lea Carlos*” lo que en su texto base aparece de manera ambigua, o combinada, como una didascalia de interlocutor prefijada al parlamento que constituye la carta:

Est. Lindo achaque. *Mar.* Tu veras
 que Elisa no puede mas,
 pues eres, Carlos, discreto.
Lea Car. Si vna mujer principal,
 porque a su padre obedece
 (Lope de Vega, fol. 88)

Como editor “en papel” (modo analógico), Madroñal confía en que el lector (humano) interpretará que la carta la lee Carlos porque así lo indica la acotación que habrá leído, aunque el texto de la carta parece corresponder tipográficamente al parlamento encabezado por “MARQUINA”. Sin embargo, como editor en TEI para ARTELOPE de la edición crítica de Madroñal, debo marcar la carta como parte de un parlamento asignado a Carlos a pesar de que no haya una didascalia de interlocutor en el texto fuente impreso. De lo contrario el cómputo de esos versos quedarían asignados a Marquina en las estadísticas. Así, mi codificación de este pasaje después de la acotación “*Lea Carlos*”, contiene el correspondiente elemento <sp> con el valor “#Carlos” en el atributo @who, y el elemento <speaker> sin englobar ningún texto.

```
<stage>Lea Carlos</stage>
<sp who="#Carlos">
  <speaker></speaker>
  <lg type="redondilla">
    <l xml:id="es-1414" n="1253"><«Si una mujer principal,</l>
```

Una codificación mecánica puede fácilmente pasar por alto que la carta la lee Carlos y dejar el contenido de la misma como parte del parlamento de Marquina, pero no quien edita el marcado de este pasaje pensando como editor de palabras como datos que tendrán su posterior procesamiento informático.

En el caso de prólogos y epílogos en determinadas obras teatrales, por ejemplo en *The Knight of the Burning Pestle* de Francis Beaumont, una edición en papel puede plasmarlos antes y después del cuerpo principal del texto sin necesidad de decidir si forman parte de este o constituyen un paratexto. La disposición tipográfica o el cambio o no de página no indicarán inequívocamente al lector qué parte estructural de la obra constituyen. Sin embargo, una edición TEI requiere decidir si se codifica como una división del texto distinta del acto o formando parte de este.

Hay ediciones en papel que incluyen una acotación de “*Aparte*” en una posición ambigua sin señalar claramente qué segmento del diálogo se dice como “aparte”. Quien edita en TEI para ARTELOPE y EMOTHE delimita ese seg-

mento en el mercado, de manera similar a como un editor crítico lo señala entre paréntesis (si esa es la convención empleada).

Una de las convenciones o normas de las bibliotecas digitales ARTELOPE y EMOTHE dicta que las didascalias de interlocutor muestren nombres completos, no abreviados. Cuando en el texto fuente hay didascalias abreviadas y estas son ambiguas, por ejemplo, “*Ser.*” interpretable como “*Servant*” or “*Sergeant*”, la edición en TEI requiere desambiguar la abreviatura, como lo haría un editor crítico cuya norma de edición contempla expandir esas abreviaturas.

Otro caso relacionado con didascalias de interlocutor ocurre cuando en el texto fuente se repite la misma didascalia en lo que, en teoría, es un solo parlamento del mismo personaje, generalmente a causa de una acotación interna que divide en dos o más partes el parlamento. Por ejemplo, en la traducción francesa a cargo de François-Victor Hugo (1866, 3: 92) de *Macbeth* de Shakespeare, en la escena quinta, la didascalia “LADY MACBETH” se repite tres veces al estar su parlamento dividido por una acotación de salida del “*serviteur*” y otra acotación de entrada de su marido. Aquí se presentan dos opciones de codificación. Una intervención mínima marcaría con tres elementos <sp> cada sección del parlamento de Lady Macbeth, cada una con sus respectivos elementos <speaker> para englobar cada didascalia “LADY MACBETH”. Esta opción reflejaría los componentes estructurales tal como están tipográficamente indicados en el texto fuente impreso. (Una variante conllevaría codificar los tres parlamentos dentro de un elemento de grupo de parlamentos <spGrp>, a su vez con su atributo @who.) Sin embargo, la lógica de la estructura dramática, no la de la disposición tipográfica del texto impreso fuente, obligaría a marcar todo con un solo elemento <sp> (y por tanto un solo elemento <speaker>), para lo que la edición digital de EMOTHE podría modificar el texto fuente mismo y eliminar las didascalias segunda y tercera, como manera de representar mejor la estructura dramática.

En ocasiones, por necesidades de adaptar el texto a las características de la plataforma de edición y las convenciones de la Biblioteca Digital, hay que recolocar una acotación. Por ejemplo, en *La corona merecida* de Lope de Vega, la edición en papel de Gómez y Cuenca (1993a: 685) que utiliza ARTELOPE muestra la acotación “[*Lea la otra carta.*]” en el margen derecho a la altura del segundo verso “y de quien vivo en su fe” del parlamento de Don Álvaro (tercer acto, verso 2373). Esta posición con el parlamento empezado es extraña y puede deberse al propósito de reflejar la posición en el manuscrito en el margen izquierdo (fol. 48 v) pero adaptándola a las convenciones de la edición en la Fundación Biblioteca Castro. Por su parte, la Biblioteca Digital de ARTELOPE solo coloca acotaciones en el margen derecho cuando son un “*aparte*”, de manera que la necesaria recolocación de la acotación obliga a decidir el lugar óptimo, al igual que hizo un editor crítico como Rodríguez-Gallego (2015: 756): la acotación se resitúa delante del parlamento de Don Álvaro.

Otra norma de las bibliotecas digitales es que ofrecen ediciones con grafía y puntuación actualizadas al uso moderno. Esto conlleva, en casos como las men-

cionadas ediciones de la Real Academia Española, a que el editor digital modernice la grafía y la puntuación, un tipo de intervención en el texto realizado en la mayoría de las ediciones críticas publicadas hoy en día.

Otra modificación importante del texto fuente es también la corrección de erratas, tal como haría un “copy editor”. Por ejemplo, como editor digital para EMOTHE de la comedia de William Congreve *Love for Love*, en la edición de G. Salgado (1968), corregí duplografías obvias como “that don’t be- | believe” o “to undergo a | a wife?”.

La codificación de la métrica en obras del teatro español ayuda a detectar deficiencias en el texto fuente. Como la edición web permite una opción para visualizar la métrica, y resultaría extraño a los lectores ver una indicación de estrofa que no se corresponde con los versos y su esquema de rima, se realizan dos tipos de intervenciones según la situación: anotación y/o modificación del texto. En *Los locos por el cielo* de Lope de Vega, para la que ARTELOPE utiliza la edición de Jesús Gómez y Paloma Cuenca para la Fundación Castro (1994), la codificación de la métrica hace sospechar algo inadecuado en la línea “caja, trompeta, bandera imperial patricio” (verso 1565), segmento hipermétrico que sigue a un serventesio que culmina una tirada de tercetos y precede a la acotación “*Vanse, y sale Maximiano y Capitan*”. Una interpretación plausible es que se trata de una impresión errónea como parte del diálogo de lo que era parte de una acotación, la cual podría tener una formulación como propone la enmienda del editor crítico Enric Bassegoda (2009: 372): “*Caja, trompeta, bandera imperial vanse, y sale Maximiano y Patricio, capitán*”. En la edición para ARTELOPE, el texto se mantiene, pero se incluye un comentario en nota.

En el caso de omisión de un verso para completar una estrofa, se inserta en el texto una nueva línea con una secuencia de puntos suspensivos entre corchetes cuadrados para indicar la laguna. Varios de esos casos se observan en *El ejemplo de casadas* de Lope de Vega, para la que ARTELOPE utiliza la edición de Jesús Gómez y Paloma Cuenca (1993b). Al marcar la redondilla se observa que falta un verso después de “¡Qué santa correspondencia!”, laguna también presente en el texto base de la *Parte XIV*. En esta misma obra, en el Baile inicial, la Fregona 1ª dice “Él se precia de cantar, | oficio que no le toca”, cuando “cantar” no rima con ninguna terminación mientras que, como ha observado Daniel Fernández (en consulta privada), “cantor” (que sería, propiamente, el “oficio” citado en el verso siguiente), rima con el “menor” y el “bebedor” de los siguientes versos (que forman parte de un villancico o canción). En la edición para ARTELOPE se enmienda a “cantor” y se advierte y justifica esta alteración en una nota: “La rima sugiere corregir a «cantor» la lectura «cantar» en la *Parte V* y en la edición de Gómez y Cuenca (p. 195).”

Un último ejemplo de intervención del editor digital tiene que ver con el paratexto de los *dramatis personae*. La comedia *El galán Castrucho* de Lope de Vega tiene la peculiaridad de que hay un elenco de figuras al principio de cada jornada. La plataforma de edición en FileMaker de ARTELOPE y de

EMOTHE no está programada para estos casos de múltiples listas intercaladas entre divisiones del cuerpo del texto. En la edición para ARTELOPE, que emplea la edición de Cotarelo para la Real Academia Española, se optó por reunir todos los personajes en una nómina inicial, y anotar la modificación. En cambio, en la edición para EMOTHE, que emplea la edición de Juan Molina para la colección CANON 60, se respetó el triple elenco y su posición, para lo que se tuvo que alterar manualmente el archivo XML-TEI con el fin de intercalar las listas donde correspondía.

Conclusión

Editar en TEI para los dos proyectos está, en primer lugar, definido por el modelo de datos, y por el software empleado (procesador de texto y gestor de base de datos FileMaker, y editor de XML oXygen), lo que permite cumplir el objetivo de conseguir una gran cantidad de textos de manera uniforme por diversos editores digitales que pueden trabajar a distancia. En segundo lugar, editar en TEI está caracterizado por el margen de intervención que tienen los responsables de la representación electrónica del texto fuente. Cuando en ARTELOPE y EMOTHE esta intervención implica alterar el texto establecido previamente (como en el caso de la modernización de la grafía, la expansión de didascalias de interlocutor, la corrección de erratas, la inserción de signos para indicar lagunas que ayudan a visualizar la métrica), la edición puede considerarse una edición “creativa” académica, según define Tanselle (1995: 13.-14) una de las variedades de edición académica, es decir, una edición que altera el texto para “mejorarlo” según un estándar de uniformidad y determinados criterios.

Esta “mejora” de los datos humanísticos (como son los textos teatrales electrónicos) también se aprecia en la labor de lo que en inglés se considera un *digital curator*, lo que en el ámbito museístico se diría en español un “conservador”, con el propósito de asegurar que la representación de los objetos de estudio funcionen bien como datos procesables e interoperables (Flanders y Muñoz). Como ha observado Susanna Allés-Torrent (2020: 64), en los últimos años, “la figura del editor ... se está alterando y dando lugar a posiciones teóricas que lo convierten en una especie de moderador”, al tiempo que evoluciona el concepto de edición digital.

Los ejemplos que he comentado en el último apartado muestran como la edición en TEI (sea un texto editado críticamente por el propio investigador o un texto previamente editado por otra persona) conlleva una labor de adaptación a las convenciones de cada proyecto, a veces de modificación del mismo texto, con decisiones editoriales y juicios interpretativos que van más allá de un marcado mecánico, de una mera descripción semántica de características y componentes estructurales del texto teatral, hasta el punto que podríamos preguntarnos si estamos ejercitando lo que se podría denominar una “codificación crítica”, en tanto que el “editor” ejercita su juicio crítico para decidir cómo interpretar y cómo re-

presentar elementos textuales, y basa sus decisiones en un conocimiento de la transmisión del texto, en la que se incluye el modelo de datos a partir del cual se representa el texto.²⁰ Si entendemos, como propone Sahle (2016: 24 y 38), que lo “crítico” o la “crítica” se refiere a las actividades que aplican un conocimiento académico y un razonamiento al proceso de reproducir o transformar un documento o texto en una edición, o cuando la edición añade información a la representación del documento para hacerlo más utilizable; si entendemos, como argumenta Allés-Torrent (2020: 73), que la “crítica” en las ediciones digitales “no se encuentra ya en la recopilación de las variantes ni en la aspiración de ofrecer un texto único, sino en la práctica y el proceso hermenéutico basado en una metodología digital (captura, estructuración, publicación)”, podemos reconocer que hay una dimensión “crítica” en lo que supone editar en TEI para ARTELOPE y EMOTHE.

20. De manera análoga, Dahlström (2010) sugiere el término “digitalización crítica” para referirse a un tipo de actividad en la digitalización en bibliotecas que tiene similitudes con la crítica textual.

Bibliografía

- “About the Folger Shakespeare”, en *The Folger Shakespeare*, [en línea] Folger Shakespeare Library, <<https://web.archive.org/web/20220318211857/https://shakespeare.folger.edu/about-the-folger-shakespeare/>>.
- ALLÉS-TORRENT, Susanna, “Crítica textual y edición digital o ¿dónde está la crítica en las ediciones digitales?”, *Studia Aurea*, XIV (2020), pp. 63-98, 20-09-23, <<https://doi.org/10.5565/rev/studiaaurea.395>>.
- BASSEGODA I PINEDA, Enric (ed.), “*Los locos por el cielo*”, en *Comedias de Lope de Vega, Parte VIII*, coord. Rafael Ramos, Lleida, Milenio, 2009, pp. 305-427.
- BESHERO-BONDAR, Elisa Eileen, “Rebuilding a Digital Frankenstein by 2018: Reflections toward a Theory of Losses and Gains in Up-Translation.” Presented at Up-Translation and Up-Transformation: Tasks, Challenges, and Solutions, Washington, DC, July 31, 2017, en *Proceedings of Up-Translation and Up-Transformation: Tasks, Challenges, and Solutions*, Balisage Series on Markup Technologies, vol. 20 (2017), 20-09-23, <<https://doi.org/10.4242/BalisageVol20.Beshero-Bondar01>>.
- BURNARD, Lou, “The TEI and XML”, en *What is the Text Encoding Initiative? How to add intelligent markup to digital resources*, new edition [online], Marseille, OpenEdition Press, 2014, 19-20-23, <<https://doi.org/10.4000/books.oep.426>>.
- CHAPMAN, Alison, y DVPP team, ed., “The Digital Victorian Periodical Poetry Schema (BETA): Schema and guidelines for encoding poems from the Digital Victorian Periodical Poetry database”, *Digital Victorian Periodical Poetry*, University of Victoria, <<https://dvpp.uvic.ca/dvpp.html>>.
- COTARELO Emilio, ed., “El galán Castrucho», en *Obras de Lope de Vega, VI: obras dramáticas*, Madrid, RAE, 1928, pp. 31-67.
- DAHLSTRÖM, Mats, “Critical Editing and Critical Digitization”, en *Text Comparison and Digital Creativity: The Production of Presence and Meaning in Digital Text Scholarship*, ed. Ernst Thoutenhoofd et al., Amsterdam, Brill, 2010, pp. 79-97.
- ESCOBAR VARELA, Miguel, *Theater as Data: Computational Journeys into Theater Research*. University of Michigan Press, 2021, <<https://doi.org/10.3998/mpub.11667458>>.
- FLANDERS, Julia, y Trevor MUÑOZ, “An Introduction to Humanities Data Curation”, *DH Curation Guide: A Community Resource Guide to Data Curation in the Digital Humanities*. Maryland Institute for Technology in the Humanities, 20-09-23, <<https://archive.mith.umd.edu/dhcuration-guide/guide.dhcuration.org/glossary/intro/index.html>>.
- FUKUSHIMA, Kailey, y Karen BOURRIER, “Inside *Digital Dinah Craik*: Feminist Pedagogy, Cognitive Apprenticeship, and the TEI”, *Journal of the Text Encoding Initiative*, XII (2019-2020), 20-09-23, <<https://journals.openedition.org/jtei/2185>>.

- FUKUSHIMA, Kailey, Karen BOURRIER, y Janice PARKER, “The Lives of Mistresses and Maids: Editing Victorian Correspondence with Genealogy, Prosopography, and the TEI”. *Digital Humanities Quarterly*, 16.1, (2022), 20-09-23, <<https://www.digitalhumanities.org/dhq/vol/16/1/000595/000595.html>>.
- GÓMEZ, Jesús, y Paloma CUENCA, ed., «La corona merecida», en *Comedias, XII / Lope de Vega*, Madrid, Turner (Biblioteca Castro), 1993a, pp. 609-705.
- , ed., «El ejemplo de casadas», en *Comedias, XV / Lope de Vega*, Madrid, Turner (Biblioteca Castro), 1993b, pp. 187-285.
- , ed., «Los locos por el cielo», en *Comedias, X / Lope de Vega*, Madrid, Turner (Biblioteca Castro), 1994, pp. 1-94.
- GONZÁLEZ-BLANCO, Elena, y José Luis RODRÍGUEZ, “ReMetCa: A Proposal for Integrating RDBMS and TEI-Verse”, *Journal of the Text Encoding Initiative*, VIII (2015), 20-09-23, <<https://journals.openedition.org/jtei/1274>>.
- Helsinki Corpus TEI XML Edition*, Helsinki: The Research Unit for Variation, Contacts and Change in English (VARIENG), University of Helsinki, 2011, <<https://helsinkicorpus.arts.gla.ac.uk>>.
- JENSTAD, Janelle, y Tracey EL HAJJ, “Converting an SGML/XML Hybrid to TEI-XML: The Case of the Internet Shakespeare Editions”, *Proceedings of Balisage: The Markup Conference 2021*, XXVI (2021), 20-09-23, <<https://doi.org/10.4242/BalisageVol26.Jenstad01>>.
- JENSTAD, Janelle, “Linked Early Modern Drama Online: A New Editorial and Encoding Platform for Shakespeare and His Contemporaries”, en *Reimagining Shakespeare Education: Teaching and Learning through Collaboration*, ed. Liam E. Samler *et al.*, Cambridge, Cambridge University Press, 2023, pp. 239-50.
- LEONHARDT, Nic, “Digital Humanities and the Performing Arts: Building Communities, Creating Knowledge”, Keynote Address SIBMAS/TLA Conference, XII, 2014, 20-09-23, <https://mappingth.hypotheses.org/files/2014/09/Nic-Leonhardt_DH-and-the-Performing-Arts_June-2014.pdf>.
- MADROÑAL, Abraham, ed., “*El ausente en el lugar*”, en *Comedias de Lope de Vega: Parte IX*, coord. Marco Presotto, Lleida, Milenio, 2007, pp. 417-531.
- MOLINA, Juan, ed., Lope de Vega, *El galán Castrucho*, en Colección CANON 60, Alicante: Biblioteca Virtual Miguel de Cervantes, 2014.
- MUÑOZ PONS, Carlos, “El proceso de edición digital en Artelope y CTCE”, *Teatro de palabras: revista de teatro áureo*, «Lope de Vega y el teatro clásico español: nuevas estrategias de conocimiento en humanidades», VII (2013), pp. 483-96.
- NILES, Rebecca, y Michael POSTON, “Re-Modeling the Edition: Creating the Corpus of Folger Digital Texts”, en *Early Modern Studies after the Digital Turn*, ed. Laura Estill *et al.*, New York / Toronto, Ier and the Arizona Center for Medieval and Renaissance Studies, 2016, pp. 117-44.
- OLEZA, Joan, “Exploring a Multilingual Digital Edition of Early Modern European Theater”, *Journal of Early Modern Cultural Studies*, 13.4 (2013), pp. 152-54.

- , “Una base de datos para una travesía compleja”, en *Renovación en el Siglo de Oro: repertorio e instrumentos de investigación*, ed. Héctor Urzáiz y M. Zubieta, Cuadernos de teatro clásico 29, Madrid, Compañía Nacional de Teatro Clásico, 2014, pp. 209-240.
- , “Los datos del TC / 12: objetivos, programas y resultados (2011-2013)”, *El patrimonio del teatro clásico español: actualidad y perspectivas. Homenaje a Francisco Ruiz Ramón*, ed. Germán García-Luengos, H. Urzaiz Tortajada y P. Conde Parrado, Valladolid, Ediciones Universidad de Valladolid, 2015, pp. 51-70.
- PORTELA, Manuel, *Literary Simulation and the Digital Humanities*. New York: Bloomsbury, 2022.
- RODRÍGUEZ-GALLEGO, Fernando, ed., “La corona merecida”, en *Comedias de Lope de Vega Parte XIV*, coord. José Enrique López Martínez, Madrid, Gredos, 2015, I, pp.583-829.
- SAHLE, Patrick, “What Is a Scholarly Digital Edition”, en *Digital Scholarly Editing: Theories and Practices*, ed. Matthew James Driscoll y Elena Pierazzo, Cambridge (UK), Open Book Publishers, 2016, pp. 19-39.
- SALGADO, Gamini, ed., “*Love for Love*”, en *William Congreve / Three Restoration Comedies*, Harmondsworth, Penguin Books, 1968, pp. 258-365.
- SHAKESPEARE, William, *Comedies, Histories and Tragedies*, London, 1623.
- TANSELLE, George Thomas, “The Varieties of Scholarly Editing”, en *Scholarly Editing: A Guide to Research*, ed. D. C. Greetham, New York, Modern Language Association of America, 1995, 9-32.
- TEI CONSORTIUM, “P5: Guidelines for Electronic Text Encoding and Interchange”, Text Encoding Initiative, 20-09-23, <http://www.tei-c.org/>.
- “Usage Guidelines for Folger Shakespeare Download Formats”, en *The Folger Shakespeare*, The Folger Shakespeare Library, 19-09-23, <<https://www.folger.edu/explore/shakespeares-works/download/usage-guidelines/>>.
- VEGA Y CARPIO, Lope de, *Doze comedias ... Novena parte*, Madrid, viuda de Alonso Martín, 1617, facsímil digital disponible en Biblioteca Digital Hispánica <<http://bdh-rd.bne.es/viewer.vm?id=0000014188&page=1>>

Por la dignidad del texto. El teatro del Siglo de Oro y de Lope de Vega en la red. Principios ecdóticos y de Humanidades Digitales¹

Ramón Valdés Gázquez

Universitat Autònoma de Barcelona

PROLOPE

ORCID: 0000-0002-7429-9709

ramon.valdes.gazquez@uab.cat

Resumen

Los textos del teatro del Siglo de Oro se han ido poniendo en la red en diferentes formatos desde imágenes de testimonios a textos procesados. En cuanto a los textos procesados, la calidad filológica es muy variada, pero han llegado a la red muchos textos no fiables establecidos con métodos y criterios caducos o sin la mínima solidez ecdótica exigible, frente al excelente panorama que ofrece hoy la edición crítica en papel. Aunque contar con una cantidad de textos importante tiene algunas ventajas, hay que lamentar que se hayan difundido y replicado gran cantidad de textos no fiables y convendrá ser prudentes en su uso para la investigación o sencillamente su lectura. Es necesario que desde el ámbito académico se publiquen y difundan las ediciones críticas y textos fiables de que ya disponemos en ediciones críticas digitales, pero también es conveniente explorar e idear estrategias de digitalización y difusión más rápida que permitan ponerlos al alcance tanto de la ciudadanía como de la comunidad investigadora según unos principios básicos e irrenunciables de Filología y de Humanidades Digitales de calidad en un contexto de Cultura y de Ciencia abiertas.

Palabras clave

Edición crítica digital; teatro del Siglo de Oro, Lope de Vega, internet, XML-TEI, Minimal Computing, Slow Science.

1. Este artículo se beneficia de la participación del autor en el proyecto de investigación “La integral dramática de Lope de Vega: textos, métodos, problemas, proyección” (PID2021-124737NB-I00), financiado por el Ministerio de Ciencia e Innovación, con fondos FEDER, así como del reconocimiento al SGR LT&T: Literary Traditions and Texts of the Early Modern Age: Italian, Iberian and Transatlantic Networks (2021SGR00155), financiado por la AGAUR, de la Generalitat de Catalunya.

Abstract

For the Dignity of the Text: Golden Age Theater and Lope de Vega Online. Editorial and Digital Humanities Principles.

The plays of the Spanish Golden Age Theater have been disseminated online in various formats, ranging from images of ancient manuscripts to processed texts. However, the Internet has seen the proliferation of numerous unreliable texts, often established with outdated methods and criteria, or lacking the necessary meticulous editorial work. This stands in stark contrast to the high standards set by contemporary critical editions in print. While the digital availability of a vast number of texts offers certain advantages, it is regrettable that these unreliable versions have been widely disseminated and replicated. Consequently, caution is advised when using them for research or casual reading. This predicament underscores the urgency of bringing critical editions and reliable texts, already available in print, to the digital realm. There is a pressing need for high-quality scholarly digital editions that adhere to the principles of Philology and Digital Humanities. Simultaneously, there is merit in exploring expeditious methods of digitalization and rapid dissemination strategies, ensuring that these critical editions and texts are promptly accessible to both citizens and the research community. Such efforts must align with the fundamental and non-negotiable principles of quality in an open culture and science scenario.

Keywords

Digital Scholarly Edition; Spanish Golden Age Theater, Lope de Vega, Internet, XML-TEI, Minimal Computing, Slow Science.

En las páginas que siguen trazaremos un panorama sobre los textos y ediciones de teatro del Siglo de Oro que se encuentran hoy en la red, con especial atención a los de Lope de Vega. Ofreceremos un balance y unas reflexiones sobre su cantidad y su calidad y, a estas alturas del siglo XXI, en un contexto práctico y teórico, una propuesta sobre estrategias y principios que conviene seguir en el futuro. Para entender mejor qué textos hay hoy, será bueno también entender cómo llegaron y se compartieron en red los primeros.

Podríamos decir que hoy la presencia del teatro del Siglo de Oro en la red, sumando diferentes formatos de presentación desde los textos procesados y publicados en HTML (muchos, en realidad, generados a partir de una codificación en XML-TEI), a los archivos PDF con imágenes de testimonios antiguos de las

bibliotecas o de ediciones críticas, o *webs* de ediciones críticas digitales propiamente dichas, es ya masiva. Según ha mantenido Germán Vega García-Luengos en diferentes ponencias y cursos la más nutrida de las tradiciones dramáticas nacionales del siglo XVII está ampliamente representada en la *web*. Si consultamos el portal *Teatro Clásico Español* de la *Biblioteca Virtual Miguel de Cervantes (BVMC)* que él coordina, a fecha de agosto de 2023, nos da un resultado de 4381 documentos entre texto en HTML, reproducciones digitales de manuscritos e impresos antiguos en formato PDF, textos codificados en XML-TEI, etc. Claro, debe considerarse que esos 4381 ítems pueden incluir hasta cinco o diez veces una misma obra en diferentes formatos y testimonios.

Pero veamos de dónde venimos para entender mejor dónde estamos y plantearnos adónde vamos o querríamos llegar, sobre todo, en términos de calidad textual.²

El teatro del Siglo de Oro en los inicios de *Internet* y la *World Wide Web*

Cabe decir que los inicios de esta digitalización —entendiendo ahora por tal cualquier iniciativa de digitalización en imagen (reproducción facsimilar) o texto— fueron muy tempranos, y llegaron fundamentalmente de la mano de dos ámbitos que no tardaron mucho en coordinarse: el de las bibliotecas y el de la academia e investigación.

Los primeros pasos del teatro del Siglo de Oro en el terreno digital se dieron por los investigadores y en forma de texto plano; formatos basados en UNIX, así como archivos creados con procesadores de textos comerciales como WordPerfect, Word y compartidos por listas de correo; o también, formatos RTF o, más adelante, PDF subidos a la red y publicados finalmente en HTML. Fueron Vern Williamsen y James T. Abraham, acompañados también por David Hildner y Matthew D. Stroud quienes compartieron generosamente los primeros textos íntegros en Internet e inmediatamente, tras su creación, en la *World Wide Web*.³ La iniciativa de Williamsen y Abraham fue más que loable y científica-

2. Aunque pretendemos mencionar los hitos principales de la llegada del teatro del Siglo de Oro en la red, no seremos exhaustivos, sino que más bien aprovecharemos la mención de algunos hitos para entender, sobre todo, cuál es la calidad textual, filológica y digital que podemos encontrar y por qué. Cabe mencionar entre los panoramas y las reflexiones previas destacables los de J. M. Lucía Megías (2008), G. Vega García-Luengos (2014) y sus “Herramientas informáticas” en la web *Casa di Lope* actualizadas por F. Antonucci, Marco Presotto (2018) y V. Casais y S. Fernández Mosquera (2022).

3. Se puede rastrear la historia en J. M. Lucía Megías (2008), J. T. Abraham (2015), Matthew D. Stroud (2015) y en la página *Comedias.org* mantenida por la Association for Hispanic Classical Theater, fuertemente relacionada con el nacimiento de este proyecto. Queda fuera de nuestra historia del teatro áureo en la red un hito fundamental en la digitalización, aunque de obligada mención: el *Teatro Español del Siglo de Oro (TESO)*, de 1997, dirigido por Carmen Simón Palmer y con asesoría de múltiples especialistas, aunque en su momento fue una iniciativa muy importan-

mente honesta. En la presentación de la colección de textos se advertía claramente: “No se debe considerar estos textos como ediciones ‘críticas’ ni definitivas”. Sin necesidad de esperar a la *web 2.0* ni que las Humanidades Digitales se bautizaran como tales en los años 2000, el proyecto se concebía de modo colaborativo: un investigador podía enviar y donar su texto, mientras otro añadía notas aclaratorias y cualquier lector enviaba enmiendas si se veían errores. Se ofrecían unos términos de uso que auguraban la filosofía de las licencias Creative Commons: libertad de utilización sin ánimo de lucro y con cita de reconocimiento.⁴ Esto favoreció la reutilización legítima por parte de otras bibliotecas virtuales, como la *Biblioteca Virtual Universal* argentina, la *IntraText Digital Library*, por otros especialistas, como Miguel Garcí-Gómez, que incluyó la colección íntegra en la suya propia (*Cibertextos interactivos*), o por los mismos investigadores en otros proyectos (Hildner reutiliza sus transcripciones para *Comedias.org* en *Teatro de los siglos de Oro*, en Québec, proyecto dirigido por R. Serrano y A. Hermenegildo). También se produjeron reutilizaciones no tan legítimas, pues hemos podido ver textos de *Comedias.org* en editoriales comerciales (todos los conocemos) que ofrecen, previo pago, servicio *print-on-demand*, sin declarar, por supuesto, el origen del texto digitalizado que toman de la red como obra de dominio público (pues la declaración sobre el uso legítimo no parecía tener la fuerza legal y coercitiva de un *copyright* formalizado al uso).

Independientemente de legitimidades, ilícitos punibles o no y éticas deontológicas, una de las ventajas indiscutibles de la red es la mayor difusión a través de la multiplicación y réplica de los textos, una vez han sido publicados. La facilidad de copia de textos digitales implica un beneficio en doble sentido: por un lado, su mayor difusión y por otro, su pervivencia. Sin embargo, si lo medimos desde una óptica ecdótica, habrá que tener muy presente también, claro, que esas ventajas se convierten en serio peligro, pues esos textos que se subieron a la red con la mejor intención, pero con declaración de provisionalidad científica, ahí quedan y, deficientemente establecidos, se reutilizan y generan su propia tradición: si los textos no se han establecido con rigor crítico, podemos estar transmitiendo, replicando y multiplicando, no solo los textos, sino los errores que, al ser provisionales, contienen. Y, siguiendo su ciclo vital inevitable en la red, esos textos, deficientemente establecidos, son base de estudios de carácter científico.

te (véase Lucía Megías 2008: 90-95). Hoy en la red, pero no en acceso abierto, es uno de esos casos en los que, algunos investigadores que ya adquirimos el costoso CD ya inoperativo, nos hemos negado a pagar dos veces para poder consultar el mismo producto que ya habíamos adquirido. No hemos podido acceder hoy al proyecto *Comedia: Spanish Drama* (Electronic Texts Center) de la Biblioteca de la Universidad de Virginia descrito por Lucía Megías (2008: 101ss.).

4. “Se les concede a los usuarios el derecho de usar o citar estos textos para cualquier fin académico y legítimo. Si se quisiera usar un texto para una producción teatral o para la publicación, se debería poner en contacto con Matthew D. Stroud para más detalles sobre los derechos y permisos.” (presentación en la página que ofrece la colección de textos).

Veamos un ejemplo de lo que venimos diciendo. En la web de *Teatro de los Siglos de Oro* de Alfredo Hermenegildo y Ricardo Serrano, de indudable mérito filológico e informático (con la revista *Teapal*, el analizador *Brocense*, debates pioneros sobre la informática textual, etc.), se ofrecen 34 comedias. De ellas, solo 9 parecen haber sido establecidas con algún criterio ecdótico y son, a decir verdad, minoría o práctica excepción los textos críticos sólidos, como la edición de Miguel Zugasti de *Todo es dar en una cosa*, de Tirso. David Hildner, que ya había transcrito una colección de textos de varios autores para *Comedias.org*, como ya hemos dicho, los recogió y los incluyó también en este proyecto y, aunque se debe valorar el mérito tecnológico de la codificación en XML así como la posibilidad de conversión y descarga en diferentes formatos, no debe olvidarse que los textos-base para las transcripciones de Hildner son escogidos sin criterio ecdótico alguno: ediciones de 1830 o 1840, de la BAE de 1873 o de principios de siglo xx de la RAE; por ejemplo, las ediciones de Leipzig de Johann Georg Keil, por lujosas que fueran y por mucho que sean buen testimonio del amor alemán por Calderón, carecen de rigor ni fundamento filológico que avale tomarlas como modelo para una edición rigurosa en nuestros tiempos. Con todo, estos textos ya electrónicos de Hildner en el nuevo proyecto de Québec son utilizados para estudios de segmentación métrica, para los que, cabe pensar, contar con textos críticos bien establecidos debería tener alguna relevancia.

Sea como sea, el teatro del Siglo de Oro empezó a tener una deseada y merecida presencia en la red gracias a los esfuerzos de estos investigadores estadounidenses y al poco nacería y tomaría el testigo la *Biblioteca Virtual Miguel de Cervantes*. De hecho, los primeros pasos de la propia *BVMC* no fueron ajenos, tampoco, a estas experiencias del lejano Oeste, y tuvo muy presente qué podía aprender de esos investigadores de las universidades de Arizona o Duke.⁵

Desde sus inicios en 1999 la *BVMC* intentó lograr el doble objetivo de ofrecer una importante cantidad de textos a la vez que una elevada calidad de los mismos. Lograr ese doble objetivo presentaba problemas cuando, además, había que conjugarlo con el imprescindible respeto de los derechos de autor (o de los editores filológicos y las casas editoriales implicadas). En algún caso la solución pasó por publicar textos con derechos de autor caducados, pero fue mala solución, pues esa caducidad implicó también caducidad científica. Aunque se crearan diferentes niveles de usuario para algunos textos (Sabido 2001: 15), al intentar satisfacer la necesidad de publicar en acceso libre una importante cantidad de textos y el deseo de que tuvieran calidad y valor científico, la *BVMC* se vio y se ha visto involucrada en alguna polémica por obras publicadas sin permiso o sin declarar con exactitud sus editores filológicos (paradójicamente, a veces —cabe

5. Según explica Stroud (2015: 2): “interest in making new plays available on our website has waned in light of the well funded *Biblioteca Virtual Miguel de Cervantes* in Spain”.

decirlo en su defensa— por el prurito de publicar buenos textos).⁶ Hay que asumir que se puede plantear una disyuntiva entre dar acceso a obras o ediciones de dominio público, sin problemas de derechos, pero con un valor científico menor o periclitado, y textos y ediciones críticas de valor filológico, pero con derechos de autor. Lo que queda fuera de lugar es entender la red como el *far West* o un *territorio sin ley*. La *BVMC* no ha sido en esto ni la única ni excepción, y son encrucijadas y polémicas en las que se han visto involucradas también bibliotecas virtuales de diferentes latitudes; de hecho, se podría decir que son conflictos comunes o propios del proceso de digitalización de muchos bienes o servicios en nuestra sociedad.⁷ Es necesaria reflexión y son necesarios equilibrios: porque un respeto a ultranza de todos los derechos implicaría una parálisis de iniciativas de Ciencia y Cultura Abiertas con que se vulnerarían los derechos de acceso del ciudadano a las mismas. Resulta significativo que las instituciones europeas estén defendiendo en ultimísimos tiempos para los datos de investigación las licencias Creative Commons-Public domain.

El portal del *Teatro Clásico Español*, TC-12, Canon 60. Testimonios, textos críticos y ediciones críticas en la red

Con todo, y volviendo a nuestro tema, tal vez podríamos decir que, en la actualidad, a través del macro-portal del *Teatro Clásico Español* de la *BVMC*, inaugurado en 2013, se han integrado más textos de este periodo y, con el diestro timón de Germán Vega, se ha avanzado por este mar proceloso. Dos factores han permitido aumentar la cantidad e interés de los textos: la digitalización masiva de testimonios antiguos por las bibliotecas y los resultados de

6. Véase a este respecto Atarés (2006), Lucía Megías (2009: 19) y (2012: 105-106) y más recientemente Fradejas Rueda (2016). Según lo relatado y demostrado por Fradejas Rueda (2016), la *BVMC* habría publicado una edición de su padre, Fradejas Lebrero, del texto medieval de Pedro López de Ayala *Libro de la caza de las aves*, diciendo que se habían basado en la edición decimonónica —y libre, por tanto, de derechos— de José Gutiérrez de la Vega, que, según declaraban, habrían cotejado y enmendado con la de Fradejas Lebrero de 1969 en Castalia. En realidad, se trataría de una mera transcripción de la de Fradejas Lebrero, según demuestra, no sin humor, Fradejas Rueda con técnicas de Lingüística Forense —y digámoslo también de Crítica Textual— rotunda y sobradamente.

7. Véase J. M. Lucía Megías (2009: 19) y (2012: 105-106), Rojas Castro (2013: 32), Faulhaber (2014: 26) así como, más allá del ámbito español, Shillingsburg (1999: 161) y Apollon, Bésile y Regnier (2014). Podríamos decir que son procesos complejos y lógicos de explotación desahogada (nunca mejor dicho), comprensión de las consecuencias y adaptación y delimitación de la posible vulneración de derechos hasta llegar a la regulación de la nueva realidad generada por la tecnología y era digital en los muy diferentes ámbitos sociales, económicos, culturales, etc. Es una de las razones por las que se habla de la *uberización* de la ciencia; se producen también en ella actuaciones competitivas que pueden resultar abusivas, desleales, incluso punibles y a veces sin mucha conciencia ni mala fe, cuando en verdad se están vulnerando derechos ajenos.

investigación de unos grupos científicos centrados en la edición crítica tan activos como generosos.

El proceso de digitalización de fondos antiguos de la Biblioteca Nacional de España comenzó en 1999, y esa digitalización tomó fuerte impulso y se hizo de manera masiva en 2008 tras la creación, en el año anterior, 2007, de la *Biblioteca Digital Hispánica*, que permitió la réplica de sus fondos en la *BVMC*. Conviene subrayar que la BNE optó por asumir, gracias a una ayuda de Telefónica, la responsabilidad de digitalizar sus propios fondos, frente a otras bibliotecas, como la de la Universidad Complutense de Madrid (o la mismísima British Library), que optó por una cooperación con Google Books, con los riesgos que comporta la cesión a una empresa privada de la gestión de un bien público.⁸

Cuando Germán Vega asumió la coordinación del portal *Teatro Clásico Español* se habían producido ya algunas felices ideas en la estructuración de la *BVMC*, como eran las páginas de “Biblioteca de Autor” o “Biblioteca de Obra”, que implicaron de por sí una potenciación de la calidad según José Manuel Lucía Megías (2009: 19 y 23-26). Germán Vega (2014) explica el proyecto en el que, sin prescindir de esos textos precedentes de la *BVMC*, basados en ediciones decimonónicas con derechos caducados o en *TESO* para el teatro del Siglo de Oro,⁹ se apuesta por incorporar, a través de convenios firmados con las distintas bibliotecas que iban digitalizando sus fondos, digitalizaciones de testimonios de interés y, gracias a convenios firmados con grupos y redes de investigación, textos críticos o incluso ediciones críticas íntegras. Digamos que con esta estrategia el compromiso entre calidad, cantidad y derechos vigentes se ha podido gestionar magistralmente. Esa estrategia, todo hay que decirlo, fue también posible en muy gran medida por el macroproyecto TC/12, vigente a través de diversos proyectos entre 2010 y 2017, y el empuje de su director Joan Oleza; ayudó, de modo general, a articular y cohesionar a los grupos de investigación y edición en Teatro del Siglo de Oro (lo que facilitó a su vez la tarea al propio G. Vega ante la *BVMC*) y acometió, de modo

8. Véase al respecto las fundamentales aportaciones de Robert Darnton (2010), que reúne un par de artículos suyos en *The New York Review of Books* de 2008 y 2009 así como la conferencia pronunciada también en este último año en la Feria del Libro de Frankfurt “Digitize, Democratize”. Una conclusión de los ensayos de Darnton, que explica con gran detalle todo el proceso de creación de Google Books, podría ser que, por elogiable y altruista que pudiera parecer —y fuera— la iniciativa de Google, al final respondía a los parámetros y objetivos de una empresa privada, y, lo que es peor, en un contexto de mercado y derechos de autor del que ningún actor se puede sustraer, se vio incluso forzada a comportarse como tal, como empresa privada, por otros agentes interesados, terceros, que, a su vez, legítimamente defendían sus derechos; al final, efectivamente, se corren serios riesgos con este modelo de negocio y cesión de la gestión del bien público (verbi-gracia digitalización de los fondos antiguos) en el que el perjudicado puede acabar siendo el propio bien público, el Estado y el ciudadano. Una perspectiva bien distinta, en defensa de la digitalización por Google, ofrecen los bibliotecarios de la UCM Magán y Tardón (2014), que no citan a Darnton, aunque sí dan fiel cuenta de la controversia.

9. Véase lo ya dicho en la nota 3; J. M. Lucía Megías denunciaba en (2009: 19) la pervivencia de textos que habían llegado a la *BVMC* con las malas prácticas del pasado.

concreto, proyectos de sumo interés como la biblioteca digital *Canon 60*, que implicó la puesta en línea de sesenta y una obras canónicas de nuestro teatro clásico en textos críticos fiables. Así, cada vez es más importante la presencia de testimonios digitalizados de fondos bibliotecarios antiguos (de interés científico *per se* en tanto que fuentes primarias), textos críticos e incluso, en el mejor y más loable de los casos, ediciones críticas digitalizadas (a destacar el caso de la *Colección Proteo* del grupo Moretianos, con 24 ediciones críticas digitalizadas, presentadas en formato PDF) con permiso y convenio con los investigadores y editores, satisfaciendo plena y simultáneamente la calidad, incluso excelencia, y vigencia científica como del respeto a los derechos de autor.¹⁰

En lo visto hasta aquí, con todo, siendo honestos, conviene admitir que, desde un punto de vista tecnológico y en términos de edición crítica digital, estos resultados pueden considerarse humildes. En los mejores casos, como *Canon 60*, los textos críticos, y por tanto filológicamente fiables, aunque muy oportunamente codificados en XML-TEI, van sin notas ni aparato crítico, por lo que no pueden considerarse como ediciones críticas plenas; o, si llevan esos aparatos, como en el caso de la *Colección Proteo*, se presentan en PDF, lo que limita enormemente su funcionalidad como edición digital (se trataría más bien de ediciones críticas *digitalizadas*, ofrecidas en PDF, un formato estático). Pero también se han producido avances importantes que combinan la excelencia filológica y tecnológica y ya se han acometido algunas que merecen plenamente el nombre de ediciones críticas digitales.

Dentro del ámbito del Teatro del Siglo de Oro cabe mencionar como ediciones críticas digitales *El perro del hortelano* de Rosa Navarro en el Centro Virtual Cervantes (ya de hace unos años, 2004, en HTML, sencilla tecnológicamente pero tal vez la pionera y muy digna edición), *La entretenida*, de Cervantes, edición de John O'Neill (2013), muy interesante proyecto del King's College por ciertas características y materiales del repositorio (por poner un ejemplo, un texto sin puntuar para su manejo y puesta en escena por compañías teatrales), así como *La traición en la amistad* de María de Zayas, dirigida por Teresa Ferrer (2015) para TEADOC (consultable hoy a través de la web de DICAT, con hiperenlaces y materiales multimedia), y *La dama boba*, dirigida por Marco Presotto (2015) y realizada en XML-TEI para PROLOPE. Esta última se convirtió en un punto de referencia en reseñas, artículos y catálogos de ediciones digitales por el concepto de archivo digital y la edición crítica y publicación en presentación sinóptica de testimonios relevantes (con sus respectivas transcripciones), como son el manuscrito autógrafo, el de "memorió" y la primera edición, relevantes todos en la historia textual de la obra, además de ofrecer, claro, el texto crítico y aparato de variantes.¹¹ En las ediciones de obras individuales

10. Ya destacaba este importante hito Marco Presotto (2018: 33-34).

11. Véase, por ejemplo, A. Rojas Castro (2017), P. Sahle (actualizado 2020) y N. Mateos Frühbeck

hay que añadir dos tesis doctorales surgidas en la Universitat de València. *La discreta enamorada*, de Lope, integrada luego en la *Biblioteca Digital EMOTHE*, de Gemma Burgos (2019) y la de Nadia Revenga (2021), una edición de *La estrella de Sevilla*, también en XML-TEI, que puso al alcance del público no solo la propia edición, descargable, sino el *software* de publicación, y que luego acabó por ofrecerse en la *Biblioteca Digital EMOTHE* complementada con tres traducciones al inglés, francés e italiano.¹²

Más allá de estas ediciones críticas digitales de obras sueltas que funcionan y son publicadas y alojadas en sus correspondientes *webs* creadas en su mayoría *ad hoc*, se han puesto en marcha otros dos proyectos que se enfrentarían a la edición de sendos corpus y ofrecerían así bibliotecas o conjuntos de ediciones críticas digitales.¹³ En primer lugar, citaremos las “Ediciones críticas” que el Grupo de Investigación Calderón ofrece en su web en la sección así llamada, donde la primera en publicarse, a inicios de 2021, fue *Las manos blancas no ofenden*, a cargo de Verónica Casais Vila. Aparte de ser una edición crítica digital plena, con cómoda visualización de variantes y notas aclaratorias, ofrece también la reproducción digital de los testimonios; carece de prólogo. Asimismo, desde otra sección de la web, la sección “Corpus”, se pueden lanzar búsquedas de cadenas de caracteres sobre 33 comedias, incluidas las tres de las que se ha realizado edición crítica digital. Con estas búsquedas se puede saber la frecuencia de uso de una palabra o acceder a sus concordancias; se ofrecen filtros muy interesantes, aparte de peculiaridades gráficas, como buscar una determinada palabra en los parlamentos de determinados personajes, o establecer horquillas temporales.¹⁴ Con todo, las “Ediciones críticas” que se ofrecen como tales en la sección de este nombre hasta el momento son solo tres.

Habíamos mencionado dos bibliotecas digitales: la otra que se está poniendo en marcha en fase beta es la biblioteca llamada *Gondomar Digital. La colección teatral del conde de Gondomar*. En ella se pretende publicar el interesantísimo corpus de la conocida como colección Gondomar que recopiló, a partir de 1595, Diego Sarmiento de Acuña, I Conde de Gondomar (1557-1626). De nuevo se trata de ediciones críticas digitales plenas, con posibilidad de visualización de variantes y notas aclaratorias y, en esta ocasión, sí se ofrecen una breve “Introducción” y una presentación más extensa de “Problemas textuales”. La

(2021). De especial interés, por ejemplo, su presentación de “variantes dinámicas”, es decir, aquellas que reflejan el proceso creativo del autor que quedó reflejado en *pentimenti* en el autógrafo.

12. Véase a estos propósitos, también, Marco Presotto (2018).

13. Más adelante nos ocuparemos de la *Biblioteca digital ARTELOPE*, que no podemos considerar aquí entre las que ofrecen ediciones críticas ni tampoco textos críticos. Tampoco es la pretensión de *EMOTHE*, aunque ocasionalmente sí haya acogido ediciones críticas y textos críticos como los ya señalados.

14. Véase Verónica Casais Vila y Santiago Fernández Mosquera (2022). Se entiende que la localización actual de las ediciones en la web del proyecto es provisional y en el futuro se ofrecerán en una biblioteca digital (383ss.); por el momento han utilizado un esquema XSD para XML y en el futuro transformarán a XML-TEI.

web ofrece también una sección de “Búsqueda avanzada” en la que se pueden buscar palabras o combinaciones, y se puede pedir la búsqueda en cualquier sección, solo en el texto o solo en las notas, lo cual resultará muy útil para futuros editores. Por el momento ofrece cinco obras en edición crítica.

En el panorama que aquí vamos presentando se puede observar que se están dando pasos importantes y se empieza a crear un corpus de textos filológicamente dignos, establecidos con rigor, e incluso de ediciones críticas digitales. No obstante, claro, la abundancia del teatro del Siglo de Oro es tal que, si somos objetivos, habremos de considerar que estos textos cuidados y estas ediciones críticas digitales son, proporcionalmente, muy escasos, por mucho que señalen en una buena dirección. Vayamos ahora al príncipe de los ingenios, paradigma de la abundancia creativa del teatro del Siglo de Oro. ¿Qué podemos decir?

Testimonios, textos y ediciones de Lope en la red

Comencemos por hacer la misma consulta que cuando hablábamos del teatro del Siglo de Oro en general. La *BVMC* nos ofrece un resultado de 640 ítems. De nuevo hay que recordar que ahí se repiten algunas obras, y que se ofrecen testimonios de todo tipo: desde digitalizaciones (reproducciones fotográficas digitales) en PDF hasta textos en HTML. Es de justicia aplaudir el esfuerzo de digitalización que están haciendo las bibliotecas, y hay que apreciar el valor científico de los testimonios digitalizados que, directamente en sus catálogos, o a través de bibliotecas digitales, ponen a nuestra disposición.

Ocuparnos ahora de Lope nos da ocasión para hacer un alto y una reflexión oportuna para poner en valor los pasos dados en el terreno de la edición crítica del teatro en coincidencia con los tiempos de la era digital. Más o menos por los años en que se creaba la *World Wide Web* y Vern Williamsen y James Abraham empezaban a compartir sus textos electrónicos en listas de correo, la gran filóloga italiana Maria Grazia Profeti, Premio Nacional de Bibliografía en 1998, reflexionaba y denunciaba respecto a la situación de la edición crítica del teatro de Lope en torno a 1990:

la tendencia más acusada es la reiterada propuesta de textos bien conocidos (por ejemplo, se continúa presentando *Fuenteovejuna*, *El caballero de Olmedo*, *El castigo sin venganza*, etc.). Lo más grave es que sigue faltando un proyecto editorial serio y programado que prevea la edición crítica del corpus dramático de Lope; se da así el caso —único, creo, en todas las literaturas nacionales europeas— de un autor fundamental que sigue circulando en ediciones desacreditadas o por lo menos que obedecen a criterios ecdóticos decimonónicos, completamente superados desde un punto de vista científico.¹⁵

15. La cita procede de Profeti (1992: 173).

Coincidiendo con esta denuncia estaba dando sus primeros pasos el grupo de investigación PROLOPE, que se había creado en 1989. Aunque es cierto que la *Parte I* de las *Comedias de Lope de Vega* no vio la luz hasta 1997, es un hecho que la edición crítica del teatro completo de Lope con el tiempo no fue solo un proyecto serio y programado, sino que ha llegado a convertirse en una realidad y un modelo de referencia, pues en 2023 podemos contar ya con la edición crítica de veintidós *partes de comedias* (las *partes*, recordémoslo, eran el formato de publicación original, género editorial que se creó con Lope, un volumen que recogía doce obras), o sea unas 250 comedias editadas por investigadores de todo el mundo tras el acopio y cotejo de todos los testimonios antiguos y el establecimiento de sus textos con riguroso método filológico, además de su anotación y estudio con los oportunos prólogos, aparatos de variantes lingüísticas, notas onomásticas, listas de erratas, etc. De modo que, respecto a lo que decía Profeti en 1992, hoy la situación y el panorama de la edición crítica de Lope (en papel, hay que recalcarlo) ha cambiado radicalmente. Y por fortuna esta reflexión se puede extender a muchos autores del teatro del Siglo de Oro, y es realmente mucho lo que se ha hecho en los últimos años, así se han abordado en diferentes proyectos de edición crítica los corpus completos de Calderón, Tirso, Pérez de Montalbán, Rojas Zorrilla, Vélez de Guevara, Mira de Amescua, Agustín Moreto, Ruiz de Alarcón, Antonio Coello, etc. Si bien es cierto, todos estos proyectos, con muy escasas excepciones, se concibieron con el objetivo de publicar las ediciones críticas, como decíamos, en volúmenes impresos, no en la red. Volvamos ahora a Lope y a la red.

¿Qué hay de Lope en la red? Podemos celebrar que en la red encontremos ya un Lope completo. Si nos fijamos en la cantidad, hay más de trescientas obras de Lope en la red, lo que nos aproximaría a su completa producción conocida. Desde los primeros pasos de la *BVMC* Lope fue un autor bien atendido. No son ediciones críticas y pueden ser transcripciones modernizadas sobre la base de, por ejemplo, los impresos antiguos que se habían transcrito antes paleográficamente en *TESO* (el ya mencionado CD-ROM *Teatro Español del Siglo de Oro*, que dirigió Carmen Simón Palmer para Chadwyck-Healey en 1998 y que ofrecía transcripción de ediciones del siglo xvii, aunque no siempre *editiones principes*). Hay que celebrar que el encargado de llevar la página de autor de Lope y la inclusión de sus obras fuera Miguel Ángel Auladell Pérez, pues, aunque no fueran ediciones críticas nos regaló ocasionalmente enmiendas muy atinadas, normalmente *ope ingenii*, aunque sin declarar a veces.¹⁶

16. <<https://www.cervantesvirtual.com/obra/biblioteca-de-lope-de-vega-0/>>, consultada el 5 de agosto de 2023. Respecto al trabajo de Auladell, puede verse, por ejemplo, el positivo balance de Daniel Fernández y Clara Monzó (2022: 1035-1036), aunque al lado de buenas enmiendas introdujera también algunos errores propios y careciera de rigor y sistematicidad metodológica (algunas enmiendas las señala, otras no).

Pero el gran proyecto que ha implicado un paso adelante gigantesco desde un punto de vista cuantitativo ha sido la *Biblioteca Digital ARTELOPE*, vinculada a la base de datos y al grupo de investigación del mismo nombre dirigido por Joan Oleza. En ella se han llegado a incluir hasta hoy nada menos que 314 textos de Lope codificados en XML-TEI, muy cerca de la producción total conservada y reconocida como auténtica. Es importante aclarar que el objetivo de estos textos era complementar la excelente base de datos *ARTELOPE*, el objetivo principal del grupo, valiosísimo complemento, claro, si se tiene en cuenta que, además de poder leer los textos que son el fundamento de su trabajo, se pueden lanzar además búsquedas de cadenas de caracteres en ellos y sobre el corpus completo (único lugar en la red y único producto digital en que esto es posible) que permiten obtener resultados de manera rápida y eficiente. Como ocurría en la *BVMC*, no nos encontramos ante ediciones críticas digitales, sino ante textos marcados en XML-TEI basados en ediciones o textos previos que son honestamente declarados en la web. Cuando accedemos al texto, en las primeras líneas, en una ficha de presentación, se especifica el texto base de la “edición digital” de la siguiente manera, por ejemplo:

Texto utilizado para esta edición digital:

Las grandezas de Alejandro. Alicante, Biblioteca Virtual Miguel de Cervantes, 1999. Edición digital a partir de: Menéndez Pelayo, Marcelino (ed.), Obras de Lope de Vega, XIV: comedias mitológicas y comedias históricas de asunto extranjero. Madrid, Atlas (BAE, CXC), 1966, pp. 330-390.

Marcación digital para Artelope:

Souto, Luz Celestina (Artelope)

Aunque en la página de inicio de la Biblioteca aparece un expositor virtual giratorio automático donde se presentan como “edición digital” (tras especificar título y autor, en este caso, “Edición digital a cargo de: Souto, Luz Celestina, Artelope”),¹⁷ de nuevo en el estilo de cita propuesto (como ocurría en la ficha de presentación: “marcación digital”) se deja muy claro, de modo absolutamente correcto, los límites de lo que es una “codificación”:

Cómo citar esta obra:

Félix Lope de Vega y Carpio. GRANDEZAS DE ALEJANDRO, Tragicomedia, LAS. Ed. Marcelino Menéndez Pelayo. Codificación digital para ARTELOPE de Luz Celestina Souto. Artelope. 2015.

<https://artelope.uv.es/biblioteca/textosAL/AL0650_LasGrandezasDeAlejandro.php>

17. Así, al menos, hasta ahora y en el momento de redacción de este artículo, 6 de agosto de 2023. Quede constancia ya que el diseño de la página de inicio de una web puede cambiar fácilmente.

Se dejan claros los límites filológicos y tecnológicos del trabajo científico acometido y la fuente o edición utilizada como base para la marcación o codificación; el valor filológico de los textos de esta Biblioteca residirá, por tanto, más en el que pudieran tener originalmente las ediciones tomadas como base que en el propio proceso de marcación (que, por supuesto, también tiene valor de por sí, pero como tal) acometido por ARTELOPE o el investigador que corresponda.¹⁸

Con los métodos de trabajo descritos tanto en la *BVMC* como en la *Biblioteca Digital ARTELOPE*, podemos decir que hay ya en la red más de 300 textos de Lope de Vega, lo que es un importante logro en términos cuantitativos. Pero ¿qué hay, entonces, de la calidad de esos textos?

Cantidad *versus* calidad

Vamos a proponer a continuación un análisis de la calidad de los textos de Lope que se han publicado en la red. Procederemos primero a un análisis de detalle confrontando algunos textos que encontramos en la red con los de sus respectivas ediciones críticas; veremos luego otros indicios de calidad y estudiaremos y analizaremos con cierto detalle el valor de las ediciones tomadas como texto base para la producción de los textos de Lope que se han publicado en la *World Wide Web*.

Así, empezaremos por la comedia *El caballero del milagro*, que editamos con Santiago Restrepo para PROLOPE. Al comparar el texto publicado en la web y nuestra edición crítica para PROLOPE se han podido detectar, en un cotejo por *loci critici*, en el de la web, en un proceso típico de transmisión textual, la reproducción pasiva de todos los errores que arrastraba o añadía la edición de los años noventa que se tomó como base hasta acumular más de treinta errores de la tradición, y sumar, en el proceso codificación y digitalización, la pérdida de tres versos íntegros en diferentes lugares de la comedia. Un cotejo exhaustivo tal vez habría revelado más errores incluso.

Podemos hacer otra comprobación rápida: si miramos los textos publicados en la red de las comedias de la *Parte XXI* que recientemente ha editado PROLOPE,¹⁹ veremos que presentan discrepancias en el número de versos, a veces con la pérdida de dos, cuatro (*Ay, verdades, que en amor...*, *La boba para los*

18. Desde un punto de vista más técnico que filológico el proceso era descrito en Carlos Muñoz Pons (2013: 485) quien proporciona un dato importante. En su proceso podían partir de un texto ya electrónico generado, por ejemplo, por la *BVMC*, gracias a un convenio firmado con esta institución. Las procedencias de sus ediciones han sido múltiples y siempre han quedado correctamente especificadas de la manera ya ejemplificada.

19. Seguimos en esto a Marco Presotto en su intervención en el congreso de la Asociación de Hispanistas alemanes en Graz en 2023 (“XXIII Deutscher Hispanistentag”, Graz, 22-25 de febrero de 2023). Ya en fase de revisión de este volumen, recibimos noticia de la publicación Presotto (2023); como ser verá, en nuestro artículo hay una importante sintonía con sus planteamientos.

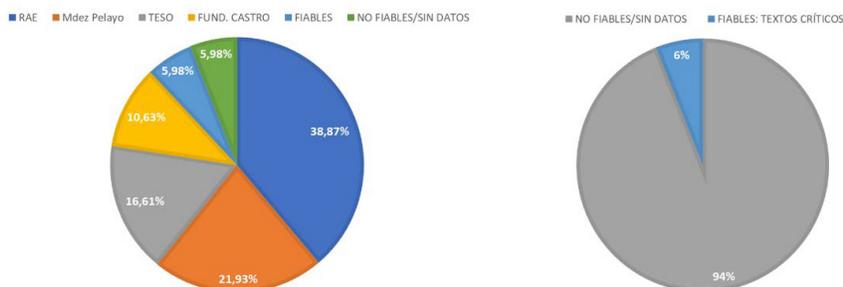
otros, *Los bandos de Sena*), y en la de *El piadoso aragonés* se pierden nada menos que unos 65; en este último caso el texto de la red, sin realizar *recensio* ni *collatio*, toma como base la *princeps* (a través de la transcripción de *TESO*) sin tener en cuenta el autógrafo que nada menos que en 1899 Menéndez Pelayo utilizó para su edición en la colección de la RAE, la cual no habría generado ningún problema de derechos de copia para su reproducción en la web (de hecho, en otras ocasiones se han seguido ediciones de Menéndez Pelayo y de la RAE). Así, de las ediciones disponibles sin derechos de copia vigentes, ni siquiera siempre se han escogido las mejores.

Algo similar pasa con *La batalla del honor*, de la *Parte VI*, cuya tradición textual conocemos y que, como *El piadoso aragonés*, dispone de autógrafo. El texto que se encuentra en la red se basa en la edición de Emilio Cotarelo para la RAE de 1917 y este, según explicaba en su prólogo, sabía de la existencia del autógrafo, pero no pudo disponer de él, de modo que el texto base que se codifica para publicar en red prescinde de un manuscrito autógrafo perfectamente conocido y localizado hace tiempo y utilizado ya en 1934 por Robert K. Spaulding y por Henryk Ziomek en 1972 para sus respectivas ediciones, además de la nuestra para PROLOPE en 2005. Por otro lado, de nuevo se produce omisión y errores de numeración que afectan a tres versos (411, totalmente omitido, y 2636-2637; además de un error de numeración a la altura del 1125) de modo que el texto crítico siempre propuesto, desde los primeros años del siglo xx, había contado 3125 versos y el que hoy se encuentra en la red cuenta 3122.

Por citar solo un caso más, podemos aludir a *El blasón de los Chaves de Villalba*, que se publicó en la *Parte X*. El texto que encontramos en la web tiene 2983 versos, mientras en la edición crítica de José Javier Rodríguez (2010) para PROLOPE son 3156. Si leemos el estudio textual de José Javier Rodríguez entenderemos perfectamente que la edición de los años noventa que se escogió y anuncia en la ficha de presentación como base para la codificación y publicación del texto la web sigue la tradición impresa deturpada por la acción de la censura, no solamente con la cancelación de esos versos, sino con la sustitución sistemática, por razones diplomáticas del momento, de cualquier alusión a Francia y los franceses por Albania y los albaneses. El manuscrito apógrafo de Miguel Sanz de Pliegos y la tradición manuscrita que permitieron a J. J. Rodríguez construir un texto crítico y recuperar los versos censurados se conoce, asimismo, desde tiempos de Menéndez Pelayo. La acción de la censura y el texto que se ofrece hoy en la web desfigura por completo una obra que Lope construyó originalmente sobre tramas principales y secundarias, amorosas y guerreras y hechos y personajes ficticios e *históricos* (de la guerra entre españoles y franceses por el reino de Nápoles) de importancia para su desarrollo y correcta comprensión. Aparte de las más de noventa alusiones concretas a Francia y franceses modificadas, entre esos 173 versos desapareció alguna escena completa. ¿Puede tal texto ofrecerse al público para leer la obra de Lope o ser la base de una investigación seria? Es el que hoy tenemos en la red.

Si, además de estos errores y deturpaciones que percibimos en obras aleatoriamente escogidas, nos fijamos en las declaraciones de procedencia de los textos que se publican (buena praxis que hay que agradecer; véase como ejemplo la remisión a la edición de Menéndez Pelayo en la ficha de ARTELOPE citada de *Las grandezas de Alejandro*; de manera similar se ofrece en *BVMC*), el panorama no mejora. En esta ocasión nos fijaremos en una muestra de 300 textos (casi la totalidad del corpus) y la información que en las propias fichas se ofrece, de modo que las cifras que manejamos son muy representativas, prácticamente descriptivas de una realidad. Atendiendo, insistimos, a la información que en las propias fichas de presentación se ofrece, veremos que más de 183 textos proceden de ediciones que, cuando menos, podríamos calificar como superadas y realizadas con criterios del siglo XIX o inicios del XX (RAE o BAE, ya fuera Menéndez Pelayo a finales del siglo XIX o inicios del XX, ya fuera la conocida como “Nueva RAE” a cargo de Emilio Cotarelo y otros); otros 50 se basan en las transcripciones electrónicas que llevó a cabo la *BVMC* de un determinado testimonio (uno solo, sin cotejo con el resto, normalmente el primer impreso en *Parte*) que de la comedia se encontrara en *TESO*; otros 32 textos se han basado en las ediciones más recientes, de los años noventa, en Biblioteca Castro, en las que los editores críticos de PROLOPE suelen detectar problemas. También hay que considerar entre los textos poco fiables y obsoletos científicamente los cinco editados por Federico Carlos Sainz de Robles para la Editorial Aguilar, sin información alguna de *recensio* ni textos base. Excepcionalmente no se ofrece dato alguno: hemos detectado seis casos. Si sumamos todo lo dicho hasta ahora, algo más de 280 textos, deberemos considerarlos como *no fiables*. Unos dieciocho textos, de los 300, quedan atestiguados como procedentes de ediciones críticas de filólogos reconocidos (de J. Badía, J. M. Díez Borque, T. Ferrer, R. Frolidí, J. Oleza, A. García Reidy, G. Serés, H. Tropé, A. Zamora Vicente). Estos textos que podemos reconocer como *fiables* (pues proceden de ediciones críticas cuyo método filológico garantiza su método y establecimiento riguroso) se reducen, entonces a un exiguo seis por ciento del total de textos presentes en la red.

ANÁLISIS DE LA PROCEDENCIA DE LOS TEXTOS DE 300 OBRAS DRAMÁTICAS DE LOPE EN LA RED



La conclusión parece firme e indiscutible: el panorama de la edición crítica y de los textos fiables del teatro de Lope en la red es desolador, muy similar, lamentablemente, al que trazaba María Grazia Profeti respecto a la edición del teatro de Lope en los años 90. Sus palabras tendrían plena vigencia si las refiriéramos al *teatro de Lope publicado en la red* a día de hoy, cuando, de hecho, como antes decíamos, el panorama de la edición crítica no solo de Lope, sino de todo el teatro del Siglo de Oro, ha cambiado radicalmente, por fortuna. Pero, sin embargo, ese cambio se ha quedado en los límites del terreno del libro (edición crítica impresa) y no ha dado el salto a la red, que se ha nutrido de ediciones de escaso rigor y caducas científicamente, como ya adelantaba muy bien Presotto (2018). Las denuncias de J. M. Lucía Megías (2009: 17-22) respecto a proyectos como los de bibliotecas virtuales “de acumulación” que priorizaban la cantidad sobre la calidad filológica han tenido poca repercusión y ese modo de hacer se ha perpetuado lamentablemente.²⁰ Lo que por ahora vamos diciendo son datos, hechos objetivos, no opiniones. Pero vayamos un paso más allá. Porque lo que aquí se está presentando como desolador, con otra luz o desde otra perspectiva podría tener su lado positivo —y también por razones objetivas—.

Calidad *versus* cantidad y velocidad para la obtención de resultados. Una ganancia

La situación se puede valorar desde distintas ópticas. Si es un hecho que el rigor y la calidad de los textos de Lope en la red dejan, en general, mucho que desear y eso es lamentable, ¿no debemos, por otro lado, celebrar como un logro, una ganancia, contar ya, aunque sea con textos llamémoslos *provisionales* en vez de *no fiables*, de casi toda su producción dramática en la red? Sin duda, no puede más que considerarse un paso adelante, una ganancia y una aportación de interés contar con ese corpus, y no solo eso, sino que, además, sobre él se puedan realizar búsquedas a través de la base de datos *ARTELOPE*.²¹ Hay que entender que no se ha pretendido realizar textos críticos, sino sencillamente ponerlos en

20. Por ejemplo: “No se justifica, en absoluto, poner en la Red ediciones ya superadas (o que nunca tuvieron un valor científico) ni que sean las únicas para dar a conocer determinados textos” (J. M. Lucía Megías, 2009: 19); aunque Lucía Megías habla, en general, de los textos de la literatura española de la *BVMC*, sus palabras, como acabamos de ver, siguen siendo perfectamente aplicables para el teatro de Lope en la red en 2023.

21. Utilizando las opciones de “Búsqueda avanzada”/“Texto de la obra”. Para lanzar búsquedas sobre el corpus completo basta no seleccionar obra alguna. No hace falta encarecer el interés que presenta la posibilidad de lanzar búsquedas sobre textos elegidos o filtrados tras la consulta de la base de datos con los criterios que se hayan estimado oportunos. Aunque las “bibliotecas de autor” (de Lope y de otros) de la *BVMC* ofrecían la posibilidad ya de generar concordancias, la cantidad de comedias de Lope sobre la que se lanza la búsqueda en *ARTELOPE*, 314 hoy (otoño de 2023), la convierte en la mejor opción.

la red y en acceso abierto, lo cual no es poco. Pero la cuestión es saber si el usuario es consciente de los límites y problemas que ello implica y si sabe que está leyendo o consultando texto científicamente *no fiable* para según qué fines.

La elaboración de ediciones críticas con rigor filológico y su puesta en línea en *webs* de calidad digital presenta una complejidad y dificultades que han llevado a la realidad que también antes comentábamos. Existen buenas ediciones críticas digitales ya disponibles en la red, pero son pocas, exigua minoría. Su elaboración implica casi la realización y diseño de *webs ad hoc*²² y resultados proporcionalmente escasos, ricos en calidad de resultados, pero pobres en cantidad. Así entra en juego un tercer factor. No conviene tener presentes cantidad y calidad, sino también velocidad.

Y es que la velocidad en la obtención y publicación de datos, en los inicios de esta era digital que hemos alcanzado, se está revelando también como uno de los factores determinantes en procesos de investigación.

La estilometría, que tantos avances relevantes y apasionantes está proporcionando en los ultimísimos años en nuestros conocimientos del teatro del Siglo de Oro tiene como punto de partida y materia prima el texto, obviamente.²³ Y de nuevo, desde esta perspectiva, hay que considerar un logro la *cantidad* de textos a disposición y la *velocidad* con que se han puesto al alcance de los investigadores. Para el proyecto *Estilometría aplicada al Teatro del Siglo de Oro (ETSO)* es fundamental alimentar a la máquina con el máximo número de textos posible. “Cuanto más numeroso sea el conjunto de piezas teatrales y dramaturgos, en general, mejores y más fiables serán estos análisis” (Cuéllar 2023: 103). Aunque siempre que han podido en *ETSO* han contado con las mejores ediciones gracias a su colaboración con distintos grupos de investigación, no siempre los textos eran los ideales (*ibidem*). Pero es que, para este objetivo parece ser poco relevante que sean textos críticos o sencillamente transcritos (incluso de modo automático con uso de *Transkribus*) de un solo testimonio no especialmente autorizado ni revisado. Para entendernos: para la estilometría es más importante contar con gran *cantidad* de textos que con textos de gran *calidad* (hablando en términos filológicos). Según explica Álvaro Cuéllar (2023: 111-112):

Afortunadamente, las pruebas estilométricas, las cuales tienen en cuenta el uso de las palabras más comunes, parecen no verse especialmente afectadas por un porcentaje irrelevante de errores en la transcripción, que suelen producirse especialmente en palabras poco comunes. Los resultados estilométricos, por tanto, son muy simi-

22. Es cierto que se está procediendo a la creación de entornos de publicación prediseñados, pero a medida que aumentamos nuestras exigencias (sin ser extravagancias), esos entornos resultan demasiado pobres y los que son más ambiciosos y adaptables, requieren ya la intervención de programadores. Véase al respecto Alvite y Rojas Castro (2022: 7-14).

23. Concretamente para Lope de Vega será de interés la consulta de Vega García-Luengos (2021) y (2023) como la página de *ETSO* y todos los avances que está realizando con Álvaro Cuéllar.

lares con obras procedentes de ediciones cuidadas y con obras procedentes de la transcripción automática. Esto quiere decir que los textos son perfectamente útiles para su análisis estilométrico.

A efectos de la estilometría, la acumulación de errores en unos treinta “tokens” sobre un total de tres mil versos (¿de cuántos “tokens” estamos hablando? unos cuantos miles más, sin duda, podemos multiplicar por seis, por siete) es por fortuna irrelevante. Recuérdese que en el proceso de trabajo mencionado se incluye la transcripción automática con herramientas o plataformas como *Transkribus*, y en esa transcripción automática se cuenta e incluye explícitamente una porción de inevitables *errores* (no ya solo los propios del testimonio utilizado, sino los añadidos por el propio *Transkribus*, por mucho que se haya entrenado y mejorado el modelo):

Las transcripciones resultantes contendrán numerosos errores y problemas, pero nos pueden ser útiles para diversos propósitos, como constituir el texto base para una edición cuidada posterior, realizar búsquedas en ellos o su procesamiento con análisis estilométricos de autoría, como es nuestro caso ahora. Estudios como el de Eder [2013] y nuestra experiencia con las pruebas informáticas nos apuntan que un porcentaje de error en las transcripciones no influye en gran medida en los análisis estilométricos, porque, además, estos suelen basarse en las palabras más frecuentes, que responden bien a los procesos automáticos de transcripción. (Cuéllar y Vega García-Luengos 2023: 135)²⁴

Efectivamente, los avances y las conclusiones de Cuéllar y Vega (y lo que en estas líneas sostienen) parecen sólidos solamente con observar cómo se agrupan coherentemente en sus diagramas o redes estilométricas los textos de los distintos autores sin importar si estos proceden de ediciones críticas, de testimonios concretos o de transcripciones obtenidas automáticamente aun con su porción

24. Creo necesario matizar que, sobre la base de mi experiencia como editor de textos, no me parece buena idea partir de un texto que lleve no solo los errores del testimonio que se digitalice, sino los de un procedimiento de transcripción automática que sabemos imperfecto. Con mucha probabilidad se acarrearán en el texto crítico, incluso después de procesos de fijación y revisión, errores de ambos (tanto del testimonio digitalizado, como los errores del proceso automático de transcripción). Me parece mejor, en cambio, una transcripción, por el momento, humana, con los errores inevitables a revisar y corregir que ello comporte, del testimonio que se prevea elegir como *texto base* para la *constitutio textus*. Una idea como la expuesta solo debe partir de un conocimiento muy profundo tanto de los principios y métodos de la Crítica Textual como de la complejidad, riesgos y problemas ecdóticos que comporta por ahora el uso de sistemas de transcripción y adquisición de textos a través de *hardware* y *software* informáticos (y de los problemas y deficiencias de todo el proceso, obtención y tratamiento tanto de la imagen con escáner como del texto con OCR y tecnologías aliadas). Incluimos en nuestra bibliografía la referencia de Eder (2013), en cuyo estudio lo que se proponía precisamente era “to verify the impact of unwanted noise in a series of experiments” (*abstract*), en línea absolutamente concordante con lo propuesto por Cuéllar, a quien agradecemos su ayuda en este tema.

de errores (aspecto al que han prestado especial atención para validar sus resultados). Es posible que basarse en ediciones críticas o limpiar, cuidar y estructurar los *messy, dirty, raw data* produjera más y mejores resultados, pero lo que ya se está consiguiendo, desde luego, resulta impresionante e inapelable (¡e imparabile!). Por otro lado, la manipulación o limpieza de los *messy, raw data* no está exenta, paradójicamente, del riesgo de incluir sesgos indeseables y perjudiciales para la propia investigación estadística.

En conclusión, en el desarrollo y avance de determinadas líneas de investigación, trabajar con textos *no fiables* no es que *pensemos que* pueda ser útil, sino que *sabemos ya positivamente* que lo es (o por decirlo con mayor verdad: no es un grave inconveniente y permite avances). Y desde esta perspectiva, el haber alcanzado la nada despreciable cifra de más de trescientas obras de Lope en la red, aunque sea en textos filológicamente no fiables o provisionales, se debe valorar como una ganancia, un logro por el que debemos dar las gracias sinceras a quienes lo han llevado a cabo y que permite realizar, entre otras, las investigaciones estilométricas que mencionamos.

Fast y Slow Science: uberización y tuitización de la ciencia en el ámbito de las Humanidades Digitales

Con todo, como prometíamos, queremos analizar la cuestión desde diferentes perspectivas, buscando un balance de ventajas e inconvenientes, y merecerá la pena hacer también otros planteamientos no tan optimistas. Tal vez el principal problema ya lo hemos expuesto: la falta de advertencia para el usuario de que los textos *no son fiables* para según qué operaciones, entre ellas una tan esperable y lógica en una biblioteca como es la lectura del texto. ¿Se le advierte de que el texto que se le ofrece no es fiable o presenta limitaciones de fiabilidad? ¿Sabe que en *El piadoso aragonés* faltan más de sesenta versos de la obra que escribió Lope? ¿Lo saben ni podrían imaginar los investigadores que no tienen una especial formación, interés ni preocupación por las cuestiones ecdóticas y que encuentran la obra al alcance de un cómodo clic ignorando su historia textual?²⁵

25. Sobre las dificultades del lector/usuario de la red para reconocer los textos de calidad, más desde el punto de vista de la *curation data*, pero también de la calidad filológica por principios y métodos establecidos por la comunidad científica, véase las reflexiones de Anne Baillot (2023): “The unsorted wealth of textual material that can be found online requires proper training for the reader or user to recognise online text quality” (p. 84); “So let us ask again: how can a reader recognise a good text from a poor one, and how can we foster access to the former and/or discourage readers from turning to the latter?” (p. 102), para definir luego como “A digital text of high quality is one that is freely available to all, in a format that makes adding further information possible, in the form of annotations, of hyperlinks, of transcriptions, of visualisations, of sonifications, and more, and which can be submitted to quality control, such as scholarly quality assessment” (p. 106); “Digital texts follow quality criteria that go beyond and add up to the editorial

Ya hemos visto cómo, una vez dado el salto a la red, los textos, en una filosofía colaborativa lícita y deseable, se van reutilizando y diseminando entre unas bibliotecas y otras (incluso libros impresos), así como para unas investigaciones u otras. El problema llega cuando, como es el caso, esos textos contienen errores. Entre los de Lope, probablemente un 94% de los que se encuentran en la red, como hemos visto. Así, uno de estos textos no fiables, que no han sido establecidos con rigor y método, con sus errores, pueden ser reproducidos y reutilizados en otras *webs* y servir de base para estudios de investigación, por ejemplo, de segmentación, donde la calidad textual es especialmente relevante, u otros muchos estudios de carácter interpretativo y hermenéutico. Parecería que el nuevo contexto digital nos obliga a una nueva defensa de la importancia de la calidad del texto y de la utilidad y necesidad de la Crítica Textual que ya habían quedado demostradas hace años, como si hubiéramos dado, en este terreno, un paso atrás. Sin embargo, esa calidad y rigor en el procedimiento depende a su vez de otro valor a la baja: la ciencia lenta (*slow science*).

En toda la parte inicial de este artículo hemos contado cómo empezaron a compartirse y difundirse en la red los primeros textos del Teatro del Siglo de Oro porque queríamos dejar claro cómo de manera inmediata, espontánea, incluso loable —ejemplo de generosidad científica e intelectual— se habían compartido incluso transcripciones provisionales preparadas con el único fin de dar clases. Es un hecho que lo que nos ofrece la tecnología se ha convertido, en cierto aspecto, en una especie de canto de sirena, de fuerte tentación que ha llevado a dar algunos pasos de una manera algo irreflexiva y precipitada sin calcular desde los posibles riesgos relativos a los derechos de autor y el reconocimiento de la autoría intelectual, hasta la calidad y valor científico mínimo, definitivo o provisional de lo que publicamos. En los inicios de la era digital esto podría parecer disculpable, pero unos treinta años después cabe reflexionar y plantearnos si podemos seguir actuando de la misma manera.

En un artículo que contiene un ataque que tildaríamos de tremendamente injusto e incluso falaz (perdón por la franqueza) contra las Humanidades Digitales por muchos conceptos, Eduard Aibar (2018) no dejó de llamar la atención sobre algunos puntos que nos deberían invitar a reflexionar.²⁶ Cómo, hasta cierto punto, nos hemos dejado seducir por lo que nos ofrecía la tecnología sin suficiente meditación ni espíritu crítico y nos hemos lanzado a la rápida publicación de resultados inmaduros o de baja calidad guiados por un afán de competencia, más que de excelencia científica. En cierto modo esta idea está en

norms that were developed over the past centuries” (p. 108); no está de más notar que plantea en idénticos términos la tensión entre cantidad y calidad en la información publicada en la red: “The tension is also one between the quantity and quality of digitised information” (p. 95).

26. Aunque Aibar no lo cita, algunos de los temas por él puestos sobre la mesa ya los había tratado Alan Liu (2012) en un artículo de referencia obligada por su visión autocrítica de las Humanidades Digitales.

la misma línea que lleva el movimiento que aboga por una *ciencia lenta*, que denuncia los errores y “las precipitaciones de la ciencia rápida” (Carlos Eduardo Maldonado, 2021).

Volvamos a plantearnos la legitimidad de los resultados cuantitativos de baja calidad en aras de alimentar a la máquina (parafraseo, no cito: “necesitamos un gran volumen de datos, o sea, muchos textos, aunque sean de baja calidad, para realizar nuestras investigaciones y esa *cantidad* de textos implicará mayor *calidad* en nuestros resultados”). Una autoridad del calibre de Christoph Schöch, conocedor profundo de las necesidades y de las dificultades de la investigación en las Humanidades Digitales, llegaba a la conclusión y hacía a la vez un llamamiento sobre la necesidad de atender no solo a los *Big data* (la estilometría maneja *Big data*)²⁷ sino también los *Smart data*. En este sentido, la urgencia de alimentar a la máquina para acometer proyectos que involucran *Big data* no se debería (aunque se pueda) hacer sobre la base de datos sucios, desestructurados, crudos, sino que también para esos proyectos, siempre será mejor contar con datos limpios, estructurados, cuidados, *fiabiles*. Por eso, Schöch (2013) propone el concepto de *Smart big data*:

we need smart big data because it can not only adequately represent a sufficient number of relevant features of humanistic objects of inquiry to enable the level of precision and nuance scholars in the humanities need, but it can also provide us with a sufficient amount of data to enable quantitative methods of inquiry that help us transgress the limitations inherent in methods based on close reading strategies. To put it in a nutshell: only smart big data enables intelligent quantitative methods.

En definitiva, y aunque valoremos como ganancia desde alguna perspectiva la cantidad de textos de Lope y otros autores a pesar de ser textos no fiables que se han procesado informáticamente y publicado en la red, de cara al futuro sería conveniente, tal vez, domar esta ciencia “desbocada”²⁸ y actuar con prudencia y responsabilidad filológica (lo es ya que los directores de *TEXORO* decidan no publicar en línea los textos transcritos automáticamente, sino que solo informe sobre su existencia y permita lanzar búsquedas sobre ellos). En la medida de lo posible, deberíamos elaborar programas escalables de trabajo que impliquen un avance ordenado y sobre fundamentos sólidos, para alcanzar *Bigger & Smarter data*. Y a la vez, y a partir de ahora, en la medida de lo posible, respetar, abogar, promover e incluso luchar por la dignidad del texto clásico en la red.

27. Una de las muchas definiciones posibles de qué son *Big data* alude a las necesidades de procesamiento de dichos datos por superordenadores, como es el caso en el proyecto *ETSO*, no solo por la estilometría, sino también por las transcripciones automáticas masivas de textos con *Transkribus* que han requerido. Téngase presente que hablamos no solo de miles de obras y autores, sino del procesamiento y comparación en múltiples planos o dimensiones de millones de versos y de “tokens”.

28. Tomo la imagen de Carlos Eduardo Maldonado (2021).

El asedio al texto y al rigor filológico en el contexto digital: el atropello de la realidad práctica y el embate teórico

Frente a los deseos recién expresados de cara al futuro, tenemos ya, como decíamos, una realidad práctica que nos ha atropellado, se nos ha echado encima, y es la que hemos descrito y valorado desde distintas perspectivas (incluso positivas) y que conviene asumir. Pero no solo nos ha atropellado esa realidad práctica, sino que podríamos hablar incluso de cierto “embate” teórico. Tanto desde la realidad práctica que implica la difusión y multiplicación de textos no fiables en la red, como, incluso, desde la teoría actual de la edición en un contexto digital parece cuestionarse la pertinencia y necesidad de la edición crítica, la exigencia de un rigor textual filológico.

Los tiempos digitales han aportado debates, planteamientos y propuestas de edición de indudable interés: desde la edición colaborativa y “social” a la “paradigmática” o la “documental”; la fluidez del texto frente a la propuesta de un único texto crítico; valoraciones tan diversas como que estamos ante una “revolución” o que solamente cambia el medio de presentación y visualización de los mismos productos y datos (las ediciones, los textos) de siempre realizados con distinta tecnología y posibilidades, sí, pero, a fin de cuentas, con el mismo método.²⁹

Así, por ejemplo, Patrick Sahle (2016), autoridad donde las haya sobre lo que es una edición académica digital, exalta su valor documental y las facilidades que ofrece para presentar el texto —y su presentación en la propia edición como pu-

29. Una de las falacias sostenidas por Eduardo Aibar (2018) es la acusación a las Humanidades Digitales de una escasa y, cuando la hay, endeble, reflexión teórica y crítica sobre la materia; siendo una falacia en lo que respecta a las Humanidades Digitales en general (basta recordar toda la web y línea de estudios *Debates in Digital Humanities*, entre otras muchísimas publicaciones que no tengo lugar para listar), también lo es y afectaría esto específicamente a los distintos ámbitos de aplicación, como el nuestro: la edición digital. Desde luego, las reflexiones teóricas sobre qué implica la digitalidad en el ámbito de la teoría y la práctica ecdótica han sido numerosísimas y muchas de extremo interés, incluyendo también aproximaciones escépticas o autocríticas. Por recordar solo a unos pocos estudiosos e hitos, y ahora sí nos detenemos, por ser nuestra materia específica, mencionaremos Charles B. Faulhaber (1991), Peter Shillingsburg (1996) y (2006); Domenico Fiormonte (2003); Hans Walter Gabler (2010); Peter Robinson (2012); el monográfico de la revista *Ecdotica* coordinado por Bárbara Bordalejo (2013), con las aportaciones de Bordalejo misma, Paul Eggert, Hans Walter Gabler, y, de nuevo, P. Robinson y P. Shillingsburg; Patrick Sahle (2013); los artículos de Andrews, Fischer, y de nuevo Robinson en el vol. X (2013) de la revista *Variants*; el colectivo de Daniel Apollon *et al.* (2014); el número monográfico del *Anuario Lope de Vega* coordinado por Ramón Valdés (2014), especialmente con las intervenciones de Paul Spence y Francesca Tomasi; Alexandra Haehnert (2015), Elena Pierazzo (2015); el colectivo de James Matthew Driscoll y Elena Pierazzo (2016), que recoge contribuciones de Marina Buzzoni y de nuevo P. Sahle; P. Robinson (2016), Dirk Van Hulle (2016), Susanna Allés-Torrent (2017), Roman Bleier *et al.* (2018), B. Bordalejo (2018), A. Rojas Castro (2018), de nuevo S. Allés-Torrent (2020), Paola Italia (2020), Baillot y Busch (2021) y M. L. Alvite y A. Rojas Castro (2022), así como las guías e informes emitidos por la MLA (2011), (2012) y (2016). Y son solo unos pocos, o poquísimos, de los que se podría y debería citar.

blicación— como un proceso fluido y una “empresa abierta” a través de sus testimonios *más que* de un texto autorizado e impuesto por el especialista. Sin embargo, ante tales propuestas, sin duda arriesgadas si se llevan a sus últimos términos, Alexandra Haehnert (2015: 14-15) advertía, recordando a Robinson y Fraistat, del peligro de convertir los proyectos de edición de las obras en meros silos de datos, imágenes o testimonios; difícilmente vastos archivos o repositorios de imágenes y transcripciones diversas serán útiles para el lector (subrayo, lector), que rara vez se acercará a la obra con un conocimiento ni curiosidad ecdótica suficientes para digerir y entender toda la información.³⁰ Esto venía a plantearse también Susanna Allés-Torrent en (2020) cuando preguntaba, provocadora, desde el título de su artículo “¿dónde está la crítica en las ediciones digitales?”. Podemos admitir que determinados contextos de investigación y difusión científica hagan ideal subrayar la fluidez del texto y el acceso a los testimonios primarios (recordemos que Patrick Sahle es historiador, no filólogo), pero es necesario reivindicar la responsabilidad del editor ante el lector que, salvo casos de versiones creadas por el propio autor o testimonios de especial interés histórico (ediciones “históricas”),³¹ debe ofrecer un texto, *uno*, lo más depurado y autorizado posible, pues no le vamos a pedir al lector que lea y conozca todos y cada uno de los testimonios de la tradición textual de *una* obra. No es su obligación ni su capacidad ni su tarea; ni siquiera, diríamos, la del investigador no interesado o formado en Crítica Textual.

Efectivamente, la edición crítica y el método riguroso de establecimiento de textos siguen siendo, en nuestra era digital, tan necesarios como siempre y en cualquiera de los casos no podremos decir que textos que han sido marcados en XML-TEI sin un rigor filológico, por mucho que sean digitales y hayan emanado en proyectos académicos son “ediciones digitales académicas”, pues (y por eso decíamos antes que parece, engañosamente, que la palabra “académica” permita una relajación del método y rigor filológico) si acudimos al propio Patrick Sahle (tan poco sospechoso de estar anclado en viejos métodos y conceptos de edición), veremos que una edición que no sea “crítica” no merecerá la categoría de “edición digital académica”. Dos de las cuatro preguntas que Sahle se planteaba al hacer su catálogo de ediciones digitales académicas, previas y definitorias por tanto del objeto y requisitos para considerarlas como tales, son las siguientes, en sus palabras:

2. Is it critical? Have rules for the processing of the material been stated and substantiated? Have these rules been applied in the light of the relevant scholarly

30. Haehnert (2015: 14-15): “vast archives of images and transcriptions are hardly useful for the reader who does not approach the work they represent with an in-depth understanding in the first place. Not that such archives and digital facsimile editions are useless —on the contrary— but they are hardly critical”.

31. Más allá del debate digital, desde una óptica puramente filológica, véase al respecto las interesantísimas aportaciones y reflexiones de Claudio Giunta (1997) sobre las “ediciones históricas”; y sobre estas, con alguna consideración digital, Valdés (2012: 414-417).

knowledge on the material, its genesis, its contexts and its reception? Does the edition add information to the representation making it more accessible, understandable and usable?

3. Is the edition of academic quality? Have the rules been applied rigorously and in a transparent manner? Are the responsibilities stated clearly? Does the edition suffice as a substitute for the previous editions or primary documents making it unnecessary to go back to them in most cases? Does it enable further scholarly research on a reliable and trustworthy basis?

Así pues, si no es sobre unas bases metodológicas ecdóticas sólidas y confiables y sobre el conocimiento y discusión de la tradición textual de una obra, no podremos considerar una edición digital —ni tanto menos un texto digital— como una “edición crítica digital”, claro, pero tampoco, porque en parte es lo mismo, *ni siquiera* una “edición digital académica” (pongo el *ni siquiera* en cursiva porque lo utilizo aquí irónicamente, frente a los que piensan que una edición digital académica tiene menos requisitos de rigor u obligaciones de método filológico). Y solo con ese método se pueden alcanzar textos y propuestas (hipótesis de trabajo) sólidos académica y científicamente. Fin de la cuestión.

¿Significa eso que solo es posible hacer ediciones digitales rigurosas según un único método, que es el nuestro? Obviamente, no seremos tan presuntuosos ni estúpidos como para mantener tal cosa. Los métodos rigurosos para realizar ediciones siempre han sido diversos y, podríamos decir, se han “ensanchado” incluso, como veíamos, en la era digital. Y pueden ofrecer un texto crítico a través del estudio de distintos testimonios, o, por el contrario, ediciones históricas o documentales (según la preferencia de denominación de Sahle) según el interés científico de esos testimonios; las ediciones podrán ser más especializadas o más divulgativas, sociales, colaborativas, dinámicas, sinópticas, paradigmáticas, etc. Y todo ello puede hacerse de manera rigurosa y tremendamente útil dependiendo de nuestro objeto de trabajo, nuestra declaración honesta de objetivos y propósitos y nuestro interés. Pero no podremos considerarlas ediciones digitales académicas si no se hacen según métodos establecidos por la Crítica Textual ni, en ningún caso, con- vendrá difundir textos deficientes o insuficientemente documentados.

Conclusiones (I).

Ilustración, defensa y propuesta de una *Biblioteca digital PROLOPE* de ediciones críticas en línea

Por supuesto, si nos dedicamos a la edición del teatro del Siglo de Oro, lo que se encuentre o no en la red a estas alturas es una responsabilidad también nuestra. Si en el grupo PROLOPE hemos editado y disponemos de más de doscientas cincuenta ediciones críticas de comedias de Lope publicadas en papel a lo largo de nuestra trayectoria de más de treinta años, es nuestra responsabilidad también ahora hacer lo posible por ponerlas en la red. Ahí cabe entonar un *mea*

culpa por el tiempo que nos está llevando lograrlo, pues desde 2012 ya nos marcamos ese objetivo, aunque ese *mea culpa* lo podemos entonar al mismo tiempo que defendemos y destacamos los logros, ya realidades, de, sobre todo, *La dama boba* dirigida por Marco Presotto (2015), pero también de la edición divulgativa y multimedia en línea de *Mujeres y criados* coordinada por Sònia Boadas y Ramón Valdés, en 2014. Sin embargo, el gran reto asumido por PROLOPE y pendiente desde 2012, obviamente, era poner en la red su producción principal, el corpus completo de las ediciones críticas de las comedias, y para ello ha diseñado la *Biblioteca digital PROLOPE*, una web que ofrecerá la oportunidad de leer en ediciones críticas solventes el corpus del teatro de Lope, además de la descarga de los archivos en diversos formatos, instrumentos de búsqueda y de análisis. Esta *Biblioteca*, cuya realización ha implicado un importante y sostenido esfuerzo en los últimos años, responde a un proceso de transformación digital (no una mera digitalización o escaneado y puesta en línea en PDF)³² a través del cual pretendemos ofrecer en acceso abierto y con sus funciones y potencialidades digitales esas ediciones críticas que hemos venido publicando en papel en toda su extensión, riqueza y complejidad, textos, prólogos, notas, aparatos, etc. En lo que venimos observando, serán ediciones más extensas y completas que las de otras bibliotecas digitales porque entendemos que el medio de publicación, la red, no obliga a una abreviación o aligeramiento (empobrecimiento, “tuitización”) de contenidos y nuestro público objetivo no va a cambiar: un lector especializado. Por contra, ofrecerlas en toda su riqueza filológica y con su funcionalidad digital implica una ganancia en términos de eficiencia en su uso con posibilidades de lectura, consulta y realización de búsquedas y descarga y reutilización por otros investigadores. Por otro lado, aunque no se conciba como proyecto divulgativo, sino de difusión, la publicación en acceso abierto también

32. Respecto a la diferencia entre “digitalización” (por influencia del inglés muy asociada a procesos de escaneado y 3D) y “transformación digital” en un contexto humanístico y de edición digital, véase Baillot (2023: sección 2), aunque en realidad el concepto de “transformación digital” es de aplicación para el reto que, en cualquier ámbito, supone la aplicación de la tecnología digital para la humanidad; la “transformación digital” implica repensar lo que el ámbito digital de publicación ofrece a nuestro objeto de estudio y procurar explotarlo en todas sus posibilidades. Digamos que, en términos de economía y comercio, una transformación digital implica exactamente lo mismo y por eso el concepto no es de aplicación solo en el ámbito de las Humanidades Digitales. En términos ecdóticos se podría hablar de la distancia que hay entre el mero escaneado de una edición en papel y todas las funcionalidades que la tecnología digital puede ofrecer en el nuevo ámbito de estudio y presentación. Por eso, como bien dice Sahle, otro de los rasgos definitorios de una edición digital es que no se puede convertir en libro (o sencillamente escanear, captar y estabilizar en un PDF) sin una pérdida fundamental e importante del valor y funcionalidades que aporta, precisamente, su tecnología digital (“A digital edition cannot be given in print without significant loss of content and functionality”, s. p.; es algo que ha costado hacer entender a los bibliotecarios en una fase en que solo nos ofrecían repositorios de documentos estables). Por supuesto, aquí son también de aplicación todas las reflexiones y debates sobre el advenimiento de lo digital en el terreno ecdótico, que ya hemos mencionado en la nota 21.

permite de por sí que se acerque el público general (y más amplio que el de la publicación en papel), que junto al especialista tendrá, como en otras bibliotecas digitales, el texto a su disposición con el apoyo de esas notas y detallados prólogos ya mencionados.³³ En la sección de “Presentación” de la Biblioteca se explican todos los principios y criterios que rigen el proyecto, y que son, fundamentalmente, los ya asumidos en un escenario de Humanidades Digitales y de Ciencia y Cultura Abiertas de uso de lenguajes de marcado difundidos (prácticamente estándares), es decir XML-TEI, cumpliendo con sus directrices y recomendaciones de buenas prácticas, así como los principios FAIR (*Findable, Accesible, Interoperable, Reutilizable*). Todo ello, sin rebajar un ápice las exigencias filológicas, sino más bien reforzándolas con nuevas fases de revisión, además de complementar el resultado con las potencialidades y funcionalidades digitales (visualización opcional e intuitiva de notas y variantes, posibilidades de búsqueda en prólogos, texto o notas, generación automática de aparato crítico, descarga de archivos en distintos formatos, hiperenlaces a testimonios, etc.).

Esto implica también todo un posicionamiento, una defensa y declaración de principios ecdóticos frente a esa realidad de una ciencia rápida que nos atropella y que inunda la red con textos poco fiables, y frente a posturas teóricas ambiguas sobre lo que significa la edición y el texto en el ámbito digital y en la red. Un ámbito que merece tanto respeto y un trabajo tan serio como el ámbito “analógico” o el formato libro (donde difícilmente se publicarían resultados provisionales o con conciencia de contener errores). Defendemos la pertinencia de la *edición crítica digital* y su difusión en la red con textos críticos filológicamente fiables, definitivos, en tanto que conclusión y plasmación de un proceso de investigación riguroso. Esto conviene aclararlo y defenderlo especialmente ahora, ante la realidad de los textos que se han publicado en la red y en momentos de cierta confusión o pérdida de respeto a nuestro principal objeto de trabajo y materia prima, el texto, y a lo que la “edición crítica digital” o “edición académica digital” implican.³⁴ Independientemente de si vamos por vías neola-

33. PROLOPE no descarta, más bien al contrario, ya está pensando y se está involucrando en acciones de investigación y transferencia dirigidas a otros públicos, pero no es este el lugar para explicarlo. Con todo, aparte de recordar la edición de *Mujeres y criados* en línea y las exposiciones (ya tres) en que se ha involucrado o ha promovido el grupo, sí podemos anunciar ya una edición social y archivo digital de *Fuenteovejuna* a cargo de David Merino Recalde (se trata de su proyecto doctoral). Para una inteligente reflexión sobre la consecuencia de la mera puesta en línea y la interactividad y adaptabilidad de las pantallas a diferentes posibilidades de lectura y públicos más allá del especialista, véase Baillot y Busch (2021).

34. En la tradición anglosajona la “scholar/scholarly edition” era una etiqueta abarcadora que permitía aludir a lo que podríamos llamar ediciones realizadas con un método riguroso, fuera el que fuera, como el de la bibliografía material, la crítica textual neolachmaniana o enfoques eclécticos. Sin embargo, ahora, en ámbito hispánico y tiempos digitales (antes nunca había tenido fortuna), se está imponiendo su uso, pero vacío de significado, como si una *edición académica* pudiera responder a muchas modalidades e inquietudes (que sí), incluso sin un requisito de rigor

chmanianas, o de la edición ecléctica, o tenemos muy presentes a la crítica genética o la bibliografía material, o las nuevas modalidades de edición surgidas con la era digital y las reflexiones que ha comportado, si realizamos ediciones críticas o de testimonios únicos por su especial interés, hay unas exigencias mínimas de rigor que nos permiten decir que hemos *editado* un *texto fiable* de una obra en un ámbito académico, científico, filológico. Solamente si procedemos a una *re-censio*, a la localización, estudio y análisis de los testimonios de la tradición de una obra y su *collatio*, su análisis y estudio, podremos realizar luego la *selectio* o la *emendatio*, la *constitutio textus* de manera rigurosa y, al menos, estos pasos son imprescindibles. Podremos adoptar o no un texto base, utilizar enfoques neolachmanianos, eclécticos u otros, realizar ediciones críticas o documentales, según la situación textual y nuestros objetivos, pero siempre unas bases teóricas y metodológicas ecdóticas y un conocimiento riguroso de la tradición textual para alcanzar y ofrecer un texto filológicamente riguroso y *fiable*.

Conclusiones (II).

Una estrategia de *Minimal Computing* para poner textos en la red compatibilizando calidad, cantidad y velocidad

Con todo, por claros que tengamos nuestros posicionamientos teóricos y el valor científico de nuestros logros, la realidad práctica, como hemos visto, insistimos, nos atropella. Como veíamos, la realización de ediciones críticas digitales, más aún, de bibliotecas o repositorios de ediciones críticas digitales resulta lenta y laboriosa. Tras varios años de trabajo, la *Biblioteca Digital PROLOPE* se publica con tres títulos, como veíamos que pasaba con las bibliotecas digitales similares en ambición de Calderón o Gondomar. Cabe esperar que, tras el costoso proceso de diseño, creación y puesta en marcha de la web, los procesos de codificación se aceleren, pero aún así, seguirá siendo un trabajo lento. Por eso, no parece que esa vía pueda revertir a corto plazo la cuestión de la calidad deplorable de los textos de Lope y la escasez de textos fiables que hoy encontramos en la red. Frente a esta urgencia, tal vez las soluciones de *Minimal Computing* ofrezcan una alternativa viable.³⁵

científico ni metodológico en el específico ámbito del establecimiento del texto (que no). Véase especialmente Paul Spence (2014) y Susanna Allés-Torrent (2020).

35. Sobre el *Minimal Computing* como una vía estratégica, abierta y accesible, altamente preservable, robusta e interoperable que cumple los principios FAIR y puede ser la respuesta más adecuada para algunos proyectos, véase Gil y Ortega (2016) así como Liviu Pop (2017) y la *web-grupo* de trabajo *Minimal Computing* con sus "Thought Pieces". Aparte de estas referencias de carácter teórico, se pueden tomar como modelos punteros que utilizan estas tecnologías, el curso *Programming Historian*, la revista *Archipelagos. A journal of Caribbean Digital Praxis*, o ediciones y colecciones textuales (como el *The CLiGS textbox*) y los proyectos de *Ediciones digitales con Minimal Computing* o el *Mini Lazarillo*.

Dado que uno de los formatos más interoperativos y que más se está utilizando (y demandando por colegas investigadores) es el .txt, además de seguir adelante con la *Biblioteca Digital PROLOPE* en el formato XML-TEI y el desarrollo de sus potencialidades digitales, nuestro grupo de investigación se ha planteado la digitalización masiva de los textos editados en simples archivos .txt, en texto plano, con el diseño de interfaces básicas para su descarga, lectura y la realización de búsquedas simples sobre cada uno de los archivos o sobre el conjunto o corpus. Tal vez el mayor reto, en esta ocasión, sea la fiabilidad y fidelidad de esa conversión en que, partiendo de los soportes en que conservamos nuestras ediciones (o bien los documentos Word en los que se enviaron a la editorial los textos, o bien PDF de maquetación, o bien el escaneado y procesamiento por OCR de nuestros libros), el texto obtenido sea absolutamente fiable, y en eso estamos poniendo todos nuestros esfuerzos iniciales. Tras la nueva revisión de los editores, podremos ofrecer nuestros textos en la web con la calidad filológica requerida, pero también satisfaciendo, esperamos, las necesidades de la cantidad y la rapidez.

Está claro que unos meros archivos .txt, aunque sumaran todos los datos (podríamos digitalizar incluso el prólogo, las notas y el aparato crítico en archivos estancos, separados; ya se verá), no serán nunca *ediciones críticas digitales*, pues la suma de los archivos .txt y los datos en archivos estancos carecerá precisamente de la funcionalidad digital y posibilidades de visualización y generación que ofrece la *Biblioteca digital PROLOPE* con sus archivos XML-TEI. Es bonito pensar hasta qué punto Gianfranco Contini estaba presagiando en 1939 lo que implica una *edición crítica digital* con su famosa simple, breve y elegante definición de lo que es una *edición crítica*: “Un’edizione critica é, come ogni atto scientifico, una mera ipotesi di lavoro, la piú soddisfacente (ossia economica) che *colleghi in sistema i dati*” (Contini, 1939, 369; subrayado mío). Ciertamente, si nosotros ofrecemos en archivos estancos los datos, pero no los relacionamos entre sí, no estamos ofreciendo, a pesar de esa suma de archivos y datos, una edición crítica. Y conviene ser conscientes de ello y de la distancia que hay entre los resultados de la *Biblioteca digital PROLOPE* y el nuevo repositorio de archivos .txt que estamos ahora concibiendo.

En este sentido, podríamos decir que los archivos .txt y su repositorio serán un producto de menor calidad digital y filológica que la *Biblioteca digital PROLOPE*, pues limitará sus funciones y posibilidades, y ofrecerá la información fragmentada. Pero al menos permitirá poner en la red *textos críticos fiables* en un formato legible con mayor velocidad y ofrecerá un producto de sumo interés y gran interoperabilidad para proyectos donde el formato de trabajo ideal sea el .txt (los hay: por ejemplo, *ETSO*). También ofrecerá, en los términos propuestos por Schöch, *Smart data*, lo cual no es poco. Porque el texto, como tal, y aunque sea desnudo, conserva todo su valor filológico y la *calidad de los datos*, pues será un texto fiable filológicamente producto de un procedimiento riguroso. Así, estaremos ofreciendo textos (*datos*) fiables a la comunidad científica y la ciudadanía para su reutilización en otros proyectos o incluso para la lectura.

Conclusiones (III). Sobre la dignidad del texto en la red

Nos hemos dilatado más de lo conveniente. Brevemente, ahora, podríamos recapitular unos principios filológicos y de Humanidades Digitales y algunas advertencias.

Desde la creación de *Internet* y de la *World Wide Web* los textos del teatro del Siglo de Oro se han ido poniendo en la red en diferentes formatos, en imágenes (una multitud de testimonios de enorme interés) y en transcripciones procesadas y publicadas en HTML. En cuanto a los textos procesados, la calidad ecdótica es variada, pero han llegado a la red demasiados textos provisionales o editados con poco rigor o con criterios incluso decimonónicos, frente al excelente panorama que ofrece hoy la edición crítica en papel; por lo que se refiere a las obras de Lope de Vega, hay que celebrar la cantidad ya puesta en línea, casi la totalidad del corpus de su producción conocida, consultable además con herramientas de búsqueda; aunque también hay que lamentar que, por razones que han quedado explicadas, se trate en su gran mayoría de textos *no fiables*. Aunque esos textos *no fiables* pueden ser útiles para realizar algunas investigaciones necesitadas de *big data*, deben utilizarse con prudencia para acometer otras o para su simple lectura. Es importante que desde el ámbito de la investigación y de los implicados en la edición digital y la difusión de textos de nuestro patrimonio literario y dramático en la red se asumieran algunas conclusiones y principios:

1. Es un logro haber alcanzado la cantidad de miles de textos del teatro del Siglo de Oro y el corpus casi completo de Lope, y podemos contar con lo hecho, *pero* con prudencia y conciencia del lector/usuario/investigador de la calidad de los textos que lee y maneja.
2. Sería deseable, en el futuro, que los textos que se ponen en acceso abierto en la red tengan calidad ecdótica e informar muy claramente al usuario y lector del trabajo filológico realizado para el establecimiento de esos textos.
3. Sería deseable que, tras la gran tarea de edición crítica que diversos grupos de investigación hemos acometido en el ámbito del teatro del Siglo de Oro, esas ediciones críticas, o al menos los textos críticos alcanzados, se pongan en línea.
4. Sería deseable no renunciar a la línea de las ediciones críticas digitales en XML-TEI, por laboriosas que resulten, pues este lenguaje de marcación es la única tecnología y estrategia suficientemente probada, robusta y extendida.³⁶

36. En nota al pie, y en el último momento, para no darle más relevancia, aunque la tiene, añadido: dadas las dificultades y la lentitud, tal vez sea bueno estar muy atentos a tecnologías de

5. En la medida de lo posible, es deseable explorar e idear estrategias de digitalización y difusión rápida de textos fiables que permitan ponerlos al alcance de la ciudadanía y la comunidad investigadora en un contexto de Cultura y Ciencia abiertas.

tratamiento del texto alternativas a la marcación XML-TEI, aunque no querría que esta advertencia se convierta en una objeción o un cuestionamiento de una decisión que por ahora parece clara: XML-TEI es por ahora la única vía posible y tecnología más difundida, robusta, interoperable por la comunidad de los editores digitales. Véanse, por ejemplo, las investigaciones de Martina Dello Buono y Francesca Tomasi (2023) basadas en tecnología CMS, HTML y Graph Data Structure. Digamos que también las ediciones críticas de Calderón se han concebido en tecnología alternativa, XML con esquema XSD, pero se marcan el objetivo, precisamente, de convertir a XML-TEI (Casais y Fernández Mosquera 2022: 378-380 y 383). No puede haber estrategia buena con diseños *ad hoc* a medida del proyecto, porque implica la renuncia a la interoperabilidad (principio básico informático y de las Humanidades Digitales). Si se descubriera o inventara una tecnología o estrategia alternativa más sencilla para la edición digital, debería extenderse tanto como la actualmente vigente entre la comunidad e incluir sistemas de conversión sencilla y fiable desde XML-TEI.

Bibliografía

- ABRAHAM, James T., "Obituary, Vern Williamsen", *Laberinto Journal*, 8 (2015), p. 1, 20-09-23, <<https://acmrspres.com/laberinto-journal-volume-8/>>.
- AIBAR PUENTES, Eduard, "La transformación neoliberal de la ciencia: el caso de las Humanidades Digitales" *ArtefaCToS Revista de estudios sobre la ciencia y la tecnología*, VII (2018), pp. 13-28, 20-09-23, <<https://doi.org/10.14201/art2018711328>>.
- ALLÉS-TORRENT, Susanna, "'Tiempos hay de acometer y tiempos de retirar': literatura áurea y edición digital", *Studia aurea: revista de literatura española y teoría literaria del Renacimiento y Siglo de Oro*, XI (2017), pp. 13-30, 20-09-23, <<https://doi.org/10.5565/rev/studiaaurea.261>>.
- ALLÉS-TORRENT, Susanna, "Crítica textual y edición digital o ¿dónde está la crítica en las ediciones digitales?", *Studia aurea: revista de literatura española y teoría literaria del Renacimiento y Siglo de Oro*, XIV (2020), pp. 63-98, 20-09-23, <<https://doi.org/10.5565/rev/studiaaurea.395>>.
- ALVITE, María Luisa y Antonio ROJAS CASTRO, "Ediciones digitales académicas: concepto, estándares de calidad y *software* de publicación", *Profesional de la información*, XXXI (2022), pp. 1-19, 20-09-23, <<https://doi.org/10.3145/epi.2022.mar.16>>.
- APOLLON, Daniel, Claire BÉLISLE y Philippe RÉGNIER (eds.), *Digital Critical Editions*, Urbana-Chicago-Springfield, University of Illinois Press, 2014.
- ATARÉS, Luisa, "Piratería editorial. Cervantes se enfrenta a los molinos en Internet", *Nueva Economía*, suplemento del diario *EL MUNDO*, número 306 (Madrid, domingo 8 de enero de 2006), 20-09-23, <<https://www.elmundo.es/suplementos/nuevaeconomia/2005/306/1136674804.html>>.
- BAILLOT, Anne, y Anna BUSCH, "Editing for Man and Machine. Digital Scholarly Editions and their Users", *Variants. The Journal of the European Society for Textual Scholarship*, XV-XVI (2021), pp. 175-187, 20-09-23, <<https://doi.org/10.4000/variants.1220>>.
- BAILLOT, Anne, *From Handwriting to Footprinting. Text and Heritage in the Age of Climate Crisis*, Cambridge, Open Book Publishers, 2023, 20-09-23, <<https://doi.org/10.11647/OBP.0355>>.
- BLEIER, Roman, Martina BÜRGERMEISTER, Helmut W. KLUG, Frederike NEUBER y Gerlinde SCHNEIDER (eds.), *Digital scholarly editions as interfaces*, Norderstedt, Books on Demand, 2018, 20-09-23, <<https://kups.ub.uni-koeln.de/9085/>>.
- BORDALEJO, Bárbara (ed.), *Work and document*, sección monográfica de la revista *Ecdotica*, X (2013), pp. 7-93.
- BORDALEJO, Bárbara, "Digital versus Analogue Textual Scholarship or the Revolution is just in the Title", *Digital Philology*, VII (2018), pp. 52-73.
- BURGOS SEGARRA, Gemma, ed., 2019: véase Vega Carpio, *La discreta enamorada*.
- CASAS VILA, Verónica, y Santiago FERNÁNDEZ MOSQUERA, "Hacia una biblioteca

- digital de las comedias de Calderón: el proceso del GIC”, *Hipogrifo*, X (2022), pp. 373-384, 20-09-23, <<https://doi.org/10.13035/H.2022.10.01.25>>.
- CERVANTES SAAVEDRA, Miguel de, *La entretenida*: véase. John O’Neill, ed., 2013.
- CONTINI, Gianfranco, “Ricordo di Joseph Bédier”, *Letteratura*, III (1939), pp. 145-152. Se cita por la edición en *Esercizi di lettura sopra autori contemporanei*, Turín, Einaudi, 1982, pp. 359-371.
- CUÉLLAR, Álvaro, “La Inteligencia Artificial al rescate del Siglo de Oro. Transcripción y modernización automática de mil trescientos impresos y manuscritos teatrales”, *Hipogrifo. Revista de literatura y cultura del Siglo de Oro*, XI (2023), pp. 101-115, 20-09-23, <<https://doi.org/10.13035/H.2023.11.01.08>>.
- CUÉLLAR, Álvaro y Germán VEGA GARCÍA-LUENGOS, “La francesa Laura. El hallazgo de una nueva comedia del Lope de Vega último”, *Anuario Lope de Vega. Texto, literatura, cultura*, XXIX (2023), pp. 131-198, 20-09-23, <<https://doi.org/10.5565/rev/anuariolopedevega.492>>.
- DARNTON, Robert, *Digitalitzar és democratitzar? El cas dels llibres*, Barcelona, Arcàdia, 2010.
- DELLO BUONO, Martina, y Francesca TOMASI, “Śivadharmā Database CMS. HTML and graph as a starting point for digital editions”, en Emanuela Carbé, G. Lo Piccolo, A. Valenti y S. Francesco (eds.), *La memoria digitale forme del testo e organizzazione della conoscenza. Atti del XII Convegno Annuale AIUCD, Siena, 5-7 giugno 2023*, Siena, Università degli Studi di Siena, 2023, 20-09-23, <http://www.aiucd.it/wp-content/uploads/2023/06/2023_aiucd_la_memoria_digitale_v1.pdf>.
- DRISCOLL, Matthew James y Elena PIERAZZO, eds., *Digital Scholarly Editing. Theories and Practices*, Cambridge, Open Book Publishers, 2016, 20-09-23, <<https://www.openbookpublishers.com/books/10.11647/obp.0095>>.
- EDER, Maciej, “Mind your corpus: systematic errors in authorship attribution” *Literary and linguistic computing*, XXVIII (2013), pp. 603-614.
- La Estrella de Sevilla*: véase Revenga (2021).
- FAULHABER, Charles B., “Textual criticism in the 21st century”, *Romance Philology*, XLV (1991), pp. 123-148.
- FAULHABER, Charles B., “*Philobiblon*. Information Technology and Medieval Spanish Literature: A Balance Sheet”, en *Humanitats a la xarxa: Món medieval. Humanities on the Web: the Medieval World*, ed. Lourdes Soriano, Helen Rovira, Marion Coderch y Gloria Sabaté, Berna, Peter Lang, 2014, pp. 15-43, 20-09-23, <<https://escholarship.org/uc/item/8kj505xd>>.
- FERRER VALLS, Teresa, ed., 2015: véase Zayas, *La traición en la amistad*.
- FIORMONTE, Domenico, *Scrittura e filologia nell’era digitale*, Turín, Bollati Boringhieri, 2003.
- FRADEJAS RUEDA, José Manuel, “Lingüística forense y crítica textual. El caso Ayala-Cervantes”, *Signa: Revista de la Asociación Española de Semiótica*, XXV (2016), pp. 193-220, 20-09-23, <<https://doi.org/10.5944/signa.vol25.2016.16927>>.

- GABLER, Hans Walter, "Theorizing the Digital Scholarly Edition", *Literature Compass*, VII (2010), pp. 43-56, 20-09-23, <<https://doi.org/10.1111/j.1741-4113.2009.00675.x>>.
- GIL, Alex, y Élika ORTEGA. "Global outlooks in digital humanities. Multilingual Practices and Minimal Computing", en *Doing Digital Humanities. Practice, Training, Research*, ed. Constance Crompton, Richard J. Lane y Ray Siemens, Londres – Nueva York, Routledge, 2016, pp. 22-33, 20-09-23, <<https://handsonhumanities.files.wordpress.com/2016/12/globaloutlooksindigitalhumanities.pdf>>.
- GIUNTA, Claudio, "Prestigio storico dei testimoni e ultima volontà dell'autore", *Anticomoderno*, III (1997), pp. 169-198, 20-09-23, <<https://heyjoe.fbk.eu/index.php/antmod>>. <www.claudiogiunta.it>.
- HÄHNERT, Alexandra, "Critical Editions and the Promise of the Digital: The Development and Limitations of Markup", Book Publishing Final Research Paper (trabajo de fin de master), repositorio PDXScholar, University Library, Portland State University, 2015. https://pdxscholar.library.pdx.edu/eng_bookpubpaper/1
- ITALIA, Paola, *Editing Duemila. Per una filologia dei testi digitali*, Roma, Salerno Editrice, 2020.
- LIU, Alan, "Where is Cultural Criticism in the Digital Humanities", en *Debates in the Digital Humanities*, ed. Matthew K. Gold, Minneapolis, University of Minnesota Press, 2012.
- LUCÍA MEGÍAS, José Manuel, "Enredando con el teatro de los Siglos de oro en la web: de los materiales actuales a las plataformas de edición", *Signa*, 17 (2008), pp. 85-129. <https://doi.org/10.5944/signa.vol17.2008.6195>
- LUCÍA MEGÍAS, José Manuel, "La edición crítica hipertextual: hacia la superación del incunable del hipertexto", en C. Castillo Martínez y J. L. Ramírez Luengo (eds.), *Lecturas y textos en el siglo XXI: Nuevos caminos en la edición textual*, Lugo, Axac, 2009, pp. 11-74.
- LUCÍA MEGÍAS, José Manuel, *Elogio del texto digital: claves para interpretar el nuevo paradigma*, Madrid, Fórcola, 2012.
- MAGÁN WALS, José Antonio, y Eugenio TARDÓN GONZÁLEZ, "Google Libros y la digitalización masiva: La aportación de la Universidad Complutense de Madrid", *Revista General de Información y Documentación*, XXIV (2014), pp. 9-24, 20-09-23, <<https://revistas.ucm.es/index.php/RGID/issue/view/2520>>.
- MATEOS FRÜHBECK, Nicolás, "Las ediciones digitales de PROLOPE, entre la edición digital y la edición digitalizada", *Revista de Humanidades Digitales*, 6 (2021), pp. 236-251, 20-09-23, <<https://doi.org/10.5944/rhd.vol.6.2021.30658>>.
- MODERN LANGUAGE ASSOCIATION, *Guidelines for Editors of Scholarly Editions*, redacción de 29-06-2011, 20-09-23, <http://mla.org/cse_guidelines>.
- , *Guidelines for Evaluating Work in Digital Humanities and Digital Media*, 2012, 20-09-23, <<https://www.mla.org/About-Us/Governance/Committees/>

- Committee-Listings/Professional-Issues/Committee-on-Information-Technology/Guidelines-for-Evaluating-Work-in-Digital-Humanities-and-Digital-Media>.
- , *Statement on the Scholarly Edition in the Digital Age*, MLA Committee on Scholarly Editions, Mayo 2016, 20-09-23, <<https://www.mla.org/content/download/52050/file/rptCSE16.pdf>>.
- MONZÓ RIBES, Clara, y Daniel FERNÁNDEZ RODRÍGUEZ, eds., Vega Carpio, Lope de, *Los bandos de Sena*, en *Comedias de Lope de Vega. Parte XXI*, coord. Gonzalo Pontón y Ramón Valdés Gázquez, Barcelona, Gredos, 2022, 2 vols, I, pp. 1019-1206.
- MUÑOZ PONS, Carlos, “El proceso de edición digital en Artelope y CTCE”, *Teatro de Palabras. Revista sobre teatro áureo*, VII (2013), pp. 483-496, 20-09-23, <https://oraprdnt.uqtr.quebec.ca/pls/public/docs/FWG/GSC/Publication/5478/122/10128/1/215131/6/O0000525071_27MunozCarlos.pdf>.
- NAVARRO DURÁN, Rosa, ed., véase Lope de Vega Carpio, *El perro del hortelano*.
- O’NEILL, John, *La entretenida by Miguel de Cervantes: A Digital, Annotated Edition and an English Translation (The Diversion)*, Londres, King’s College London, 2013, 20-09-23, <<https://entretendida.outofthewings.org/about-project/index.html>>.
- PIERAZZO, Elena. *Digital Scholarly Editing: Theories, Models and Methods*, Farnham (Surrey), Ashgate, 2015.
- POP, Liviu, “A low te(a)ch approach to Digital Humanities”, *Digitalia*, LXII (2017), pp. 83-88, 20-09-23, <<https://doi.org/10.24193/subbdigitalia.2017.1.06>>.
- PRESOTTO, Marco (ed.), 2015: véase Vega Carpio, *La dama boba*.
- PRESOTTO, Marco, “La edición crítica digital del teatro del Siglo de Oro”, *Cuadernos de AISPI*, XI (2018), pp. 31-46, 20-09-23, <<https://dx.doi.org/10.14672/0.2018.1426>>.
- PRESOTTO, Marco, “Ecdótica y Humanidades Digitales en el estudio del Teatro clásico español: una experiencia”, *Cuadernos Revista de Humanidades Digitales*, VIII (2023), 2-03-24, <<https://doi.org/10.5944/rhd.vol.8.2023.37267>>.
- PROFETI, Maria Grazia, “La obra dramática de Lope de Vega”, en Francisco Rico, coord., *Historia y crítica de la Literatura Española*, vol. III/1, primer suplemento: *Siglos de Oro: Barroco*, coord. por Aurora Egido, Editorial Crítica, Barcelona, 1992, pp. 172-184.
- REVENGA, Nadia, ed., *La estrella de Sevilla*, edición crítica digital (en *La Estrella de Sevilla y las potencialidades de la edición digital*, tesis doctoral dir. por Teresa Ferrer Valls) y multilingüe (con las traducciones de Henry Thomas, 1935, al inglés; Eugène Baret, 1869, al francés, y Gherardo Marone, 1969, al italiano), Biblioteca Digital EMOTHE, Universitat de València, Valencia, 2021, 20-09-23, <https://mobiroderic.uv.es/bitstream/handle/10550/79924/tesis_nrevenga_EstrelladeSevilla_EdicionDigital.pdf;sequence=1&isAllowed=y>. <<https://emothe.uv.es/biblioteca/editions.php?n=4>>.

- ROBINSON, Peter, "Towards a Theory of Digital Editions", *Variants*, X (2013), pp. 105-131.
- ROBINSON, Peter, "The Digital Revolution in Scholarly Editing", en *Ars Edendi Lecture Series*, vol. IV, eds. B. Crostini, G. Iversen y B. M. Jensen, Estocolmo, Stockholm University Press, vol. IV, pp. 181-207, 20-09-23, <<http://dx.doi.org/10.16993/baj.h>>.
- RODRÍGUEZ RODRÍGUEZ, José Javier, ed., Lope de Vega Carpio, *El blasón de los Chaves de Villalba*, en Lope de Vega Carpio, *Comedias de Lope de Vega. Parte X*, coord. Ramón Valdés Gázquez y María Morrás, Lérida, Milenio-Universitat Autònoma de Barcelona, 2010, tomo III, pp. 1187-1361.
- ROJAS CASTRO, Antonio, "Lope de Vega's *La Dama Boba*. Critical edition and digital archive", *RIDE – A review journal for digital editions and resources*, V (2017), 20-09-23, <<https://ride.i-d-e.de/issues/issue-5/damaboba/>>. <<https://doi.org/10.18716/ride.a.5.3>>.
- ROJAS CASTRO, Antonio, "La edición académica digital. De las teorías del texto a la visualización de la información", en eds. Isabel Galina-Russell, José Francisco Barrón-Tovar, Miriam Peña-Pimentel, David Domínguez-Herbón, Ernesto Priani-Saisó, Adriana Álvarez-Sánchez, *Humanidades digitales: edición, literatura y arte*, México D.F., Bonilla Artigas, 2018, pp. 47-80, 20-09-23, <<https://doi.org/10.17613/tef7-bx87>>.
- ROSSELLI DEL TURCO, Roberto, «After the Editing is Done: Designing a Graphic User Interface for Digital Editions», *Digital Medievalist*, VII (2011), 20-09-23, <<http://doi.org/10.16995/dm.30>>.
- SABIDO, Vicente, "La Biblioteca Virtual Miguel de Cervantes", *El Profesional de la Información*, X (2001), pp. 15-18. DOI: 10.1076/epri.10.11.15.6773
- SAHLE, Patrick, *Digitale Editionsformen: Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, Norderstedt, Books on Demand, 3 volúmenes, 2013.
- SAHLE, Patrick, *A Catalog of Digital Scholarly Editions*, v. 3.0, snapshot 2008ff, compiled by , última actualización del 28/09/2020; 20-09-23, <<https://v3.digitale-edition.de/>>.
- SCHÖCH, Christoph, "Big? Smart? Clean? Messy? Data in the Humanities", *Journal of Digital Humanities*, II (2013), sin número de página, 20-09-23, <<https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>>. <<https://zenodo.org/record/8432>>.
- SHILLINGSBURG, Peter, *Scholarly Editing in the Computer Age: Theory and Practice*, Ann Arbor, University of Michigan Press, 1996.
- SHILLINGSBURG, Peter, *From Gutenberg to Google: Electronic Representations of Literary Texts*, Cambridge, Cambridge University Press, 2006.
- SIMÓN PALMER, Carmen (dir.), *Teatro Español del Siglo de Oro. Base de datos a texto completo [TESO]*, versión 3.0, Chadwyck-Healey España, Madrid, 1998. Actualmente en línea (véanse Recursos).
- SPENCE, Paul, "Edición académica en la era digital: modelos, difusión y proceso

- de investigación”, *Anuario Lope de Vega. Texto. Literatura. Cultura*, XX (2014), págs. 47-83.
- STROUD, Matthew D., “Obituary, Vern Williamsen”, *Laberinto Journal*, VIII (2015), pp. 2-3, 20-09-23, <https://acmrs.asu.edu/sites/default/files/2020-01/laberinto_v8.pdf>.
- VALDÉS, Ramón, “*Los sueños* de Quevedo: variantes, versiones y estrategias de edición”, en Eugenia Fosalba y Gonzalo Pontón, eds., *La escondida senda: estudios en homenaje a Alberto Blecuá*, Madrid, Castalia, 2012, pp. 383-421.
- VALDÉS, Ramón (ed.), *Edición digital y edición crítica*, sección monográfica de la revista *Anuario Lope de Vega. Texto, literatura, cultura*, XX (2014), pp. 1-121, 20-09-23, <<https://revistes.uab.cat/anuariolopedevega/issue/view/v20>>.
- VAN HULLE, Dirk, “Modelling a Digital Scholarly Edition for Genetic Criticism: A Rapprochement”, *Variants. The Journal of the European Society for Textual Scholarship*, XII-XIII (2016), pp. 34-56, 20-09-23, <<https://doi.org/10.4000/variants.293>>.
- Variants. The Journal of the European Society for Textual Scholarship*, X (2013).
- VEGA CARPIO, Lope de, *La dama boba*, edición crítica y archivo digital de Marco Presotto con colaboración de Sònia Boadas, Eugenio Maggi y Aurèlia Pessarrodona, Barcelona – Bolonia, PROLOPE – Alma Mater Studiorum – Università di Bologna, Centro di Risorsse Multimediali, 2015. DOI: 10.6092/UNIBO/LADAMABOBA
- VEGA CARPIO, Lope de [autoría discutida], *La estrella de Sevilla*: véase Nadia Revenga, ed.
- VEGA CARPIO, Lope de, *La discreta enamorada*, edición crítica digital de Gemma Burgos Sagarra, Biblioteca digital ARTELOPE – Colección EMOTHE, Valencia, Universitat de València, 2019, 20-09-23, <<https://emothe.uv.es/biblioteca/editions.php?n=1>>.
- VEGA CARPIO, Lope de, *Mujeres y criados*, edición en línea coordinada por Sònia Boadas y Ramón Valdés, Biblioteca virtual PROLOPE, Bellaterra, Universitat Autònoma de Barcelona, 2014, 27-01-24, <<https://prolope.uab.cat/obras/index.html>>
- VEGA CARPIO, Lope de, *El perro del hortelano*, edición de Rosa Navarro Durán, Centro Virtual Cervantes (Obras de referencia: Clásicos Hispánicos), Madrid, Instituto Cervantes, 2004, 20-09-23, <https://cvc.cervantes.es/obref/perro_hortelano/default.htm>.
- VEGA GARCÍA-LUENGOS, Germán, “Planos y planes de un escenario privilegiado para el teatro clásico español: el macroportal TCE de la Cervantes Virtual”, *Cuadernos de Teatro Clásico*, número monográfico editado por Héctor Urzáiz y Mar Zubieta, *Renovación en el Siglo de Oro: repertorio e instrumentos de investigación*, XXIX (2014), pp. 369-394.
- VEGA GARCÍA-LUENGOS, Germán, “Las comedias de Lope de Vega: confirmaciones de autoría y nuevas atribuciones desde la estilometría (I)”, *Talía. Revista de Estudios Teatrales*, III (2021), pp. 91-108, 20-09-23, <<https://doi.org/10.5209/tret.74625>>.

VEGA GARCÍA-LUENGOS, Germán, “Para la delimitación del repertorio de comedias auténticas de Lope: confirmaciones de autoría y nuevas atribuciones desde la estilometría (II)”, *Anuario Lope de Vega. Texto. Literatura. Cultura*, XXIX (2023), pp. 469-544, 20-09-23, <<https://doi.org/10.5565/rev/anuariolopedevega.477>>.

ZAYAS, María de, *La traición en la amistad*, edición a cargo del grupo TEADOC dirigida por Teresa Ferrer Valls, Valencia, Universitat de València, 2015. <<https://dicat.uv.es/latraicionenlaamistad>>.

WEBS, EDICIONES, RECURSOS Y PROYECTOS CITADOS

Archipelagos. A journal of Caribbean Digital Praxis, 20-09-23, <<https://archipelagosjournal.org/>>.

Biblioteca Digital ARTELOPE, 20-09-23, <<https://artelope.uv.es/biblioteca/>>.

Biblioteca Digital PROLOPE, (de próxima publicación), <<https://prolope.uab.cat/biblioteca>>.

Biblioteca Virtual Miguel de Cervantes, 20-09-23, <<http://www.cervantesvirtual.com/>>.

Biblioteca Virtual Universal, 20-09-23, <<https://www.biblioteca.org.ar/>>.

Canon 60, 27-01-24, <<https://www.cervantesvirtual.com/partes/682072/canon-60-la-coleccion-esencial-del-tc12-teatro-clasico-espanol>>.

Cibertextos interactivos, 20-09-23, <<https://people.duke.edu/~garci/cybertexts/>>.

Comedias.org (The Association for Hispanic Classical Theater), 20-09-23, <<http://www.wordpress.comedias.org/>>

Ediciones digitales con Minimal Computing, 20-09-23, <<https://raffazizzi.gitlab.io/minimaldigipub/es/proyectos/>>.

Estilometría aplicada al Teatro del Siglo de Oro (ETSO), 27-01-24, <<https://etso.es/>>.

The CLiGS textbox, 20-09-23, <<https://github.com/cligs/textbox>>.

Gondomar Digital. La colección teatral del conde de Gondomar, 20-09-23,
<<http://gondomar.tespasiglodeoro.it/>>.

Grupo de Investigación Calderón de la Barca: Ediciones críticas, 27-01,24,
<https://www.calderondelabarca.org/ediciones_criticas>.

Intratext Digital Library, 20-09-23, <<http://www.intratext.com/>>.

Mini Lazarillo, edición mínima del Lazarillo de Tormes, 20-09-23,
<<https://minilazarillo.github.io/>>.

Minimal Computing, “Thought pieces”, 20-09-23,
<<http://go-dh.github.io/mincomp/thoughts/>>.

Programming Historian, 20-09-23, <<https://programminghistorian.org/>>.

TC/12 Red del Patrimonio Teatral Clásico Español, 27-01-24,
<<https://tc12.uv.es>>.

Teatro Clásico Español, portal de la *Biblioteca Virtual Miguel de Cervantes*, 20-09-23,
<https://www.cervantesvirtual.com/portales/teatro_clasico_espanol/>.

Teatro Español del Siglo de Oro (TESO), 20-09-23,
<<https://www.proquest.com/legacyredirect/teso>>. Vid. Simón Palmer (1998).

Teatro de los siglos de Oro, 20-09-23,
<<https://www.uqtr.ca/teatro/entrada/t.html>>.



Editar una obra y los testimonios de su recepción. El caso de Góngora

Mercedes Blanco

Sorbonne Université
citrine.blanco@gmail.com

Aude Plagnard

Université Paul-Valéry, Montpellier 3
aude.plagnard@gmail.com
ORCID 0000-0001-5107-4921

Resumen

Reflexionamos acerca del trabajo realizado en Πólemos, un proyecto de humanidades digitales que se propone la edición y estudio de los documentos relativos a la recepción de Góngora en su siglo. Mostramos que el objeto mismo se prestaba a un tratamiento digital, lo que nos lleva a una reflexión sobre la naturaleza de la edición digital. Esta no es una transformación optimizada de la edición filológica, sino una creación de objetos nuevos a partir de materiales textuales diversos, lo que ha llevado a hablar de “nuevo paradigma” o “paradigma digital” opuesto al “paradigma tipográfico”. Con la experiencia limitada de Πólemos y con un par de ejemplos más, podemos intuir que estos objetos ponen de manifiesto la historicidad de los textos, tanto en su génesis (tradicción material), como en su interpretación y valoración históricamente variables (recepción). Este acercamiento se completa en la segunda parte con una consideración detenida de un aspecto técnico tan importante como el lenguaje XML-TEI, y el modo en que aprendimos a utilizarlo para conjugar la edición misma, bien definida por Sahle como una representación crítica de documentos históricos, junto con los medios de su comprensión y contextualización. Por otro lado, describimos formas de lectura distante de este corpus derivadas del etiquetado TEI, como la construcción de herramientas que permiten visualizar redes de relaciones entre textos y entre agentes.

Palabras clave

Edición digital; Proyecto Πólemos; etiquetado TEI; minería de datos; análisis de redes.

Abstract

Editing a Work and the Documents of its Reception. The Case of Góngora.

We reflect upon the work undertaken in Πólemos, a Digital Humanities project focused on the editing and analysis of the Gongorine polemic, specifically delving into the documents of Góngora's reception in his century. In the first part, we elucidate the rationale behind the selection of this subject for digital treatment, prompting a consideration of the essence of digital editing, which is not merely an optimized transformation of philological editing, but rather a creation of novel entities derived from the textual sources. Drawing insights from the limited experience of Πólemos and a few other examples, we surmise that these entities compel us to emphasize the historical dimension of texts, both in their genesis (material tradition) and in their subsequent reception and interpretation. This perspective is further developed in the second part, with a detailed consideration of such an important technical aspect, as the XML-TEI language. We expound upon our acquisition of proficiency in using this language, elucidating how it serves as a tool for intertwining the act of editing—defined by Sahle as a critical representation of historical documents—with the production of means for comprehending and contextualizing these documents. We delineate diverse methods for constructing a distant reading of this corpus through TEI encoding, including such the development of tools that facilitate the visualization of networks of relationships among texts and agents.

Keywords

Digital scholarly editing; Πólemos project; TEI markup; data mining; network analysis; data mining.

Este artículo se propone dos cometidos distintos, aunque conectados: el primero es dar una idea del proyecto Πólemos,¹ edición colectiva, filológica y digital de la obra completa de Góngora y de los numerosos testimonios de su accidentada y polémica recepción en el siglo xvii; el segundo consiste en explicar el tipo de

1. <https://obvil.sorbonne-universite.fr/projets/gongora-et-les-querelles-litteraires-de-la-rennaissance>

análisis específico que las herramientas digitales, y especialmente el lenguaje de marcado de la Textual Encoding Initiative,² nos permitieron desarrollar. En el marco de este proyecto, nuestra relación con la edición digital se desarrolló de una forma más bien empírica. Si algo podemos aportar a la empresa colectiva que supone este monográfico, será partiendo de la reflexión sobre una práctica limitada a este proyecto y no de un aparato teórico a priori, cuya adquisición se superpuso a nuestra vocación y formación de filólogas, críticas e historiadoras. Formulamos, a guisa de preámbulo, ciertas observaciones sobre la naturaleza misma de la edición digital, tal como la hemos practicado y como hemos observado que funciona en otros proyectos, relativos también a la literatura del Siglo de Oro, aun a riesgo de que sean evidencias o truismos para quienes están mejor armados en esta materia.

I. Reflexión sobre la naturaleza de la edición digital mediante un par de ejemplos

Partimos de la constatación siguiente: la edición digital tal vez no sea igualmente relevante y oportuna, o en todo caso igualmente necesaria, para cualquier texto o cualquier corpus. Existen obras literarias que han circulado y circulan en ediciones destinadas a la imprenta y al papel, las cuales cumplen perfectamente con sus objetivos, tanto científicos, como de divulgación y difusión. En cambio, se dan objetos textuales para los cuales la edición digital es en cierto modo insustituible: aunque en teoría podrían editarse en papel, en la práctica son inabarcables por este método tradicional, por razones económicas y también lógicas, derivadas de su carácter múltiple, disperso, complejo o evolutivo —este es el caso del corpus gongorino, que editamos en Πólemos. Para decirlo de otro modo, y aunque esto no suponga novedad para quienes conocen bien este campo, la edición digital no es solo una manera de interpretar, codificar y difundir un objeto textual, ni una simple alternativa u optimización de un tratamiento tradicionalmente filológico.

Es más, no solo el método para realizar y presentar la edición de un objeto textual (texto o corpus) *X* cambia si sometemos *X* a un tratamiento digital, produciendo una oposición entre su edición filológica tradicional (EFT) y su edición filológica digital (EFD),³ sino que es el mismo objeto el que cambia: en vez de una oposición simple entre EFT(*X*) y EFD(*X*) —la edición filológica tradicional de *X* contra su edición digital—, tenemos una oposición en dos dimen-

2. La Text Encoding Initiative (TEI) es un consorcio que desarrolla y mantiene de manera colaborativa un estándar web para la representación de textos en formato digital, tal y como se especifica en su página web, <https://tei-c.org/>.

3. Hay quien, mejor que de “edición filológica” prefiere hablar de “edición académica”, traduciendo “scholarly edition”, como muy bien explica Susanna Allés-Torrent (2020: 90), pero confesamos que encontramos algo nebuloso el concepto.

siones entre EFT(X) y EFD (X^d), siendo X^d una nueva versión del texto inicial y una especie de mutante, un avatar de X en el mundo digital. Por ejemplo, la edición digital de *La dama boba* de Lope por Marco Presotto⁴ no es, obviamente, una variante de la edición en papel de la misma comedia, con un traje un poco distinto (pantalla en lugar de papel, notas que pueden verse no a pie de página sino pulsando el número de nota, etc.). Al pasar de una a otra, cambia la naturaleza del objeto editado.⁵

Como se sabe, la edición filológica tradicional, destinada al impreso o, lo que es lo mismo, a establecer un patrón destinado a la reproducción en tantos ejemplares idénticos como se desee, tiende a contemplar el texto como una unidad, y a convertirlo idealmente en una entidad invariable y definitiva. Claro que se trata de una tendencia, y que en la práctica existen tantas ediciones impresas distintas como editores de un texto dado. Con esa salvedad, cada persona que emprende un trabajo filológico de edición destinada al impreso, llevada por la lógica y tradición de su labor, supone a la obra literaria una identidad única y se propone ofrecer un texto óptimo, el que más se acerque a la voluntad definitiva del autor. Se consignan las variantes, fruto, por lo común, de errores o de interferencias debidos a accidentes de transmisión, o bien residuos del proceso de elaboración del texto conservados en versiones primitivas o borradores. Pero el llamado ‘aparato crítico’ es relegado a una posición marginal, y casi solo sirve de garantía científica, puesto que permite comprobar a quien lo desee que la tarea de cotejar los testimonios y de basar en ellos el texto ha sido ejecutada de acuerdo con las reglas del arte (aunque no haya un consenso perfecto sobre dichas reglas).

La edición digital, en cambio, incluso cuando tiene por punto de partida una edición crítica hecha según esas famosas reglas del arte, como sucede en el caso del mencionado trabajo del equipo de Marco Presotto, la presenta no rara vez junto con un archivo de documentos representativos de la denominada ‘tradicón textual’, lo que convierte el texto en una entidad plural, un haz de textos sin jerarquía unívoca. En otras palabras, el dispositivo editorial impone, o al menos sugiere, la idea de que no hay una *Dama boba* de Lope de Vega sino varias o muchas: 1. la del autógrafo original digitalizado, 2. la del mismo autógrafo transcrito, 3. la del impreso de la comedia en una de las *Partes (princeps)* cuya impresión tuvo a cargo Lope, digitalizado también, 4. la de una copia manuscrita

4. <http://damaboba.unibo.it/>.

5. Para todo esto, y lo que sigue, véase Allés-Torrent, 2020, que expone en una brillante síntesis el conjunto de las ideas y debates sobre la naturaleza de la edición digital a partir de principios de siglo hasta la fecha del artículo. Resume en estos términos el modo en que la edición digital se hace abogada de lo múltiple frente a lo unitario: “La idea de texto, en los últimos decenios, ha ido evolucionando, y se ha cuestionado la cultura del texto ‘único’ abogando por su fluidez, el valor de la variante y su dimensión social. Así también lo ha hecho el concepto de edición digital, vista hoy como una especie de archivo digital donde tienen cabida todos aquellos materiales relacionados con el texto” (p. 64).

ta llamada M que se considera derivada de la memorización por un espectador, digitalizada, 5. la de la transcripción de esta copia, 6. y, por fin, la de la edición moderna dicha ‘crítica’.⁶ Aunque en este trabajo, con toda su prudencia y seriedad italianas, esa posibilidad no se contemple, podría imaginarse la introducción en este dispositivo editorial de las reescrituras de la comedia para montajes teatrales de hoy, como la que ofrece Juan Mayorga (Molanes 2011). Esta multiplicidad es consecuencia de una atención a los aspectos materiales de la tradición, que sin los medios digitales no habría sido posible restituir al lector de forma tan exhaustiva, colocando además cada estado del texto en un mismo nivel.⁷ Como escribe Marco Presotto:

La edición no anotada, o hasta la edición crítica tradicional con sus exigencias y limitaciones también derivadas del soporte en papel, por lo común reduce mucho la atención a todos estos aspectos materiales de la tradición textual, relegándolos generalmente a un apéndice o a una síntesis introductiva.⁸

El objeto editado tiende, pues, a difractarse en varias facetas, varios documentos, y posiblemente esta multiplicidad tenga su correlato en la pluralidad de investigadores que colaboran para establecer la edición, un trabajo de equipo donde el responsable principal, Marco Presotto, actúa respaldado nada menos que por tres colegas: Sònia Boadas, Eugenio Maggi y Aurèlia Pessarrodona. No solo es posible consultar estos múltiples avatares de *La Dama boba* desde la misma página web, con solo escoger uno de los hipervínculos o bifurcaciones que aparecen en la portada y en las páginas siguientes, sino que se facilita al “usuario” la posibilidad de leerlas en paralelo, como una partitura polifónica, escogiendo entre varias combinaciones (por ejemplo, edición crítica junto al autógrafo, o el autógrafo junto al impreso o su transcripción).⁹ La metáfora musical

6. Así se explica al respecto Marco Presotto en la presentación del proyecto editorial: “Dada la importancia de los testimonios de *La dama boba* y del diferente ámbito de realización (O = manuscrito autógrafo; A = edición autorizada por el autor; M = copia derivada supuestamente de un proceso de memorización ilegal), el proyecto se ha focalizado en la transcripción de los documentos y en el objetivo de ofrecer tales transcripciones de manera sinóptica, para poderlas comparar entre ellas y con la edición crítica” (<http://damaboba.unibo.it/proyecto>).

7. De esta idea, Patrick Sahle ofrece una expresión particularmente enérgica: “Of course, the content of digital editions can—in theory—be printed out. And, of course, the text of digital editions could still be read on paper. However, a main characteristic of a digital edition is its representation of a potentially large number of documents in a potentially limitless number of different views, such as facsimile, diplomatic transcription and reading versions” (Sahle 2016: 27).

8. Marco Presotto, “¿Por qué una edición digital?”, <http://damaboba.unibo.it/proyecto>.

9. “El usuario tendrá así a disposición la edición crítica y los tres testimonios transcritos, que podrá comparar entre ellos gracias a una organización de la pantalla repartida en módulos distintos (hasta 3) para dar cabida a las distintas selecciones elegidas. Asimismo, las opciones de búsqueda podrán referirse al texto crítico pero también a todos los testimonios o a la selección de documentos que se consideren oportunos” (<http://damaboba.unibo.it/proyecto>).

tiene desde luego un valor meramente heurístico, y es tan inadecuada como sugerente, puesto que al lector y espectador interesado por la intriga dramática y por los personajes y sensible a la poesía que se desprende de esta bonita comedia no le es posible, a fin de cuentas, escuchar más que una línea melódica a la vez, y la interferencia de las demás impediría el pleno disfrute de la gracia y el ingenio de la obra: en cambio para el estudioso, este despliegue de versiones diferentes tiene la virtud de exhibir el carácter poliédrico del texto, lo que produce un valor añadido, cognitivo y práctico, innegable.¹⁰

En este caso, parecerá quizá abusivo hablar de un objeto textual X convertido en X^d, o sea en un objeto que es digital por esencia y no solo por el modo de su presentación, puesto que nada impide, incluso con los medios tradicionales, leer en paralelo varios testimonios de un texto: de hecho, esa lectura comparada o cotejo es la condición previa del trabajo crítico del filólogo, la primera etapa del procesamiento ecdótico. Pero se trata de una especie de andamio destinado a desaparecer o a ser relegado al desván o la bodega del “aparato crítico”, a beneficio del resultado final, la edición crítica tal como se concibe al menos desde el humanismo y la imprenta (aunque el término y la formalización de la técnica aparecieron en el siglo XIX, los principios esenciales son muy anteriores). En cambio, en ediciones digitales como la que tomamos por ejemplo, y a la que se ha dado, de hecho, un valor de modelo, puesto que ha sido muy citada y reseñada, la distancia entre lector y editor tiende a esfumarse, al menos a perder nitidez, invitando al “usuario” a un papel activo aunque ocasional de filólogo aficionado. Y es que, de modo bastante lógico, para quienes entran como receptores en esta clase de edición, se habla de “usuario” más que de lector. La lectura, tal como se entiende desde la era de la imprenta, es solo una parte, y quizá no la más importante del uso de estas ediciones, que más que tales, se llaman a sí mismas proyectos. Es decir que el usuario tiende a consultar los textos, a saltar de un punto a otro más que a leerlos, o sea recorrerlos linealmente del principio al fin; o, si los lee, es asunto suyo, pero no exactamente lo que prevé el dispositivo.

Más radical es la novedad del objeto producido por el trabajo editorial en un proyecto como *PRESOLO. Prácticas editoriales y sociabilidad literaria*, alojado en la universidad de Córdoba, y debido a un equipo formado por Ignacio García Aguilar, Pedro Ruiz Pérez, Rafael Bonilla Cerezo y Victoria Aranda Arribas, además de colaboradores ocasionales.¹¹ PRESOLO tiene, según propia de-

10. Como indica Allés-Torrent, 2020, p. 72, este tipo de edición que suma todos los testimonios ha sido realizado en el caso de obras fundamentales, como el *Quijote* y la *Celestina*, o con textos medievales, como en el *Corpus electrónico de manuscritos del cancionero castellano del siglo XV*, dirigido por Dorothy Severin. En general, a estos testimonios, no se suma una edición crítica, como sucede con *La dama boba*. Otro caso excepcional, que asocia edición crítica y transcripción de los testimonios es el de Luís de Góngora, *Soledades*. Edición crítica digital, ed. Antonio Rojas Castro, <<http://soledades.uni-koeln.de/>>.

11. <https://www.uco.es/presolo/>

claración, tres objetivos. El primero es “recopilar y editar todos los paratextos compuestos por el Fénix”, junto con los que redactaron otros autores para obras de Lope de Vega, lo que da lugar a una “biblioteca de paratextos”: un conjunto muy amplio de textos breves (543 documentos, según García Aguilar 2023) consultable en línea. El segundo objetivo perseguido es el marcado TEI del corpus, ideado para realizar búsquedas combinadas (y no solo por palabras sueltas, como en un texto desprovisto de marcado conceptual) atendiendo al sentido “pragmático, literario y discursivo de todas las piezas”. Naturalmente tanto ese marcado como el buscador se basan en ciertas hipótesis previas, puesto que hacen hincapié en el sentido “pragmático” de los textos. O sea, contemplan estos textos como palabras en acción destinadas a revestir la figura del autor de ciertos valores, construyendo lo que llaman estos investigadores “identidad autorial” y haciendo visibles “las tensiones y vínculos más o menos clientelares que influyeron en la actividad literaria del Fénix”. Sin embargo, el carácter sesgado de la interpretación así sugerida al usuario de esta página web queda compensado por la índole abierta y provisional del proyecto, compartida por todos los de este tipo, que pueden verse como llamadas (pocas veces atendidas, bien es verdad) a una discusión y una crítica constructiva. El tercer objetivo consiste en “la visualización” mediante el software Gephi,¹² de las relaciones bilaterales o reticulares en las cuales se inserta Lope de Vega. Nótese que el concepto mismo puede verse como una proyección del funcionamiento actual de las llamadas redes sociales en la web de la era digital.

Baste para nuestro presente objetivo mencionar que, en este caso de PRE-SOLO, se construye, bajo el nombre de biblioteca, un conjunto de documentos literarios y administrativos —puesto que entre ellos figuran no solo los prólogos, dedicatorias y aprobaciones sino también las licencias y tasas—, que no existió como tal históricamente. Por supuesto, se puede recorrer con cierta fruición esta biblioteca de paratextos al modo de herbolario de ejercicios retóricos (dados el ingenio y la destreza en el manejo del castellano de Lope y de los demás autores de muchas de estas piezas) pero su propósito y naturaleza son típicamente científicos: reunir documentos como materiales de un protocolo experimental, la llamada “búsqueda avanzada”, para poner a prueba ciertas hipótesis de sociología literaria. Así concebido, lo que en PRESOLO se propone no es un texto X que hubiera podido editarse por medios tradicionales: lo que se edita es un objeto X^d, un objeto de la era digital, irrealizable y tal vez inconcebible fuera de ella. Otro ejemplo, más limitado en lo que pretende abarcar, y quizá, por ello mismo, más convincente (de momento) de este tipo de método se produce cuando se someten textos teatrales a un marcado destinado a cartografiar las interacciones entre los personajes: ¿quién habla, qué número de versos se adjudican a cada personaje, a quién habla, durante cuantos versos y en qué momen-

12. <https://gephi.org/>

to de la obra? La visualización de los datos a modo de redes (con nodos o vértices, y enlaces o aristas) mediante el programa Gephi puede permitir, en casos ideales, comprobar una hipótesis de gran alcance sobre la poética y el proceso de escritura de un autor.¹³ Es lo que hicimos nosotros en el teatro de Góngora en el seno del proyecto Πόlemos, pero sin interpretar los resultados:

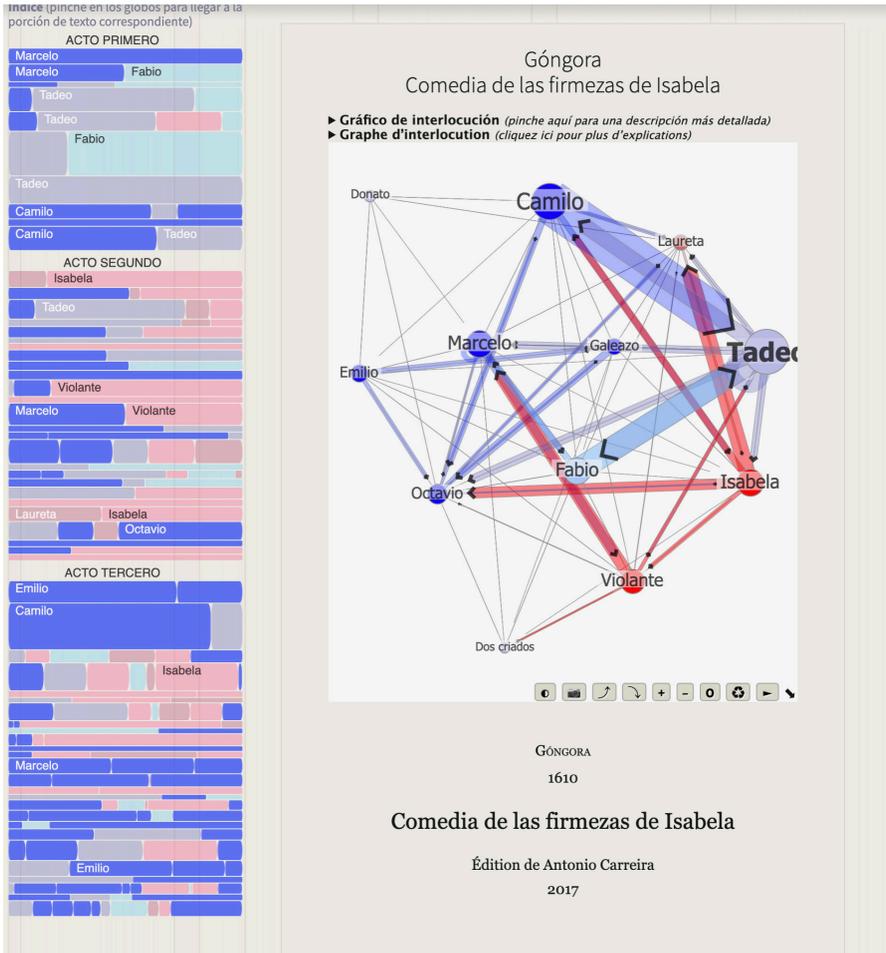


Fig. 1. Gráfico de interlocución en la comedia *Las firmezas de Isabela* (proyecto Πόlemos).¹⁴

13. Véase, como ejemplo brillante, Carmona, 2021: la comparación de las redes de interacción en dos comedias muy afines temáticamente, *El mayordomo de la duquesa de Amalfi* y *El perro del hortelano*, permite a la autora demostrar que Lope “escribe con falsilla” o sea que la segunda de las comedias es en cierto modo una reescritura de la primera, aunque nadie lo diría fijándose únicamente en los textos, y solo se comprueba al hacer aparecer la estructura subyacente.

14. https://obvil.sorbonne-universite.fr/corpus/gongora/gongora_firmezasisabela.

Como se comprobará, en nuestro proyecto, Góngora también tiende a difractarse y multiplicarse y, por ello mismo, la naturaleza digital de la edición influye no solo en nuestra manera de proceder, sino también en los postulados iniciales y los resultados de la investigación.

II. El proyecto **Πólemos**: Góngora con los testimonios de su recepción

II.1 *Objetivos y definición del corpus*

El proyecto Góngora del Labex OBVIL, que de modo informal también llamamos Πólemos, se asigna por objetivo editar el conjunto de los testimonios de la polémica gongorina. Nació en 2014 y aunque ha dado lugar a muchísimo trabajo y a resultados que creemos interesantes, dista de estar terminado. En este caso, la idea surgió de consideraciones filológicas y hermenéuticas. Si bien en su inicio, razones de oportunidades institucionales y de curiosidad por nuestra parte nos llevaron a adoptar la forma de ediciones digitales, alojadas en un macroproyecto de investigación de humanidades digitales literarias, llamado Labex OBVIL (Laboratoire d'Excellence «Observatoire de la Vie Littéraire»),¹⁵ sin embargo, esta dimensión digital fue determinante y, en algunos aspectos, imprescindible para alcanzar el resultado provisional al que hoy hemos llegado.

Antes de señalar brevemente las mencionadas consideraciones filológicas y hermenéuticas que nos llevaron a diseñar este proyecto editorial, es necesario aclarar lo que entendemos por testimonios de la polémica gongorina. No nos propusimos editar tan solo los escritos cuyo objeto único o principal fue atacar a Góngora o defenderlo frente a quienes lo atacaban, sino el conjunto de los testimonios de su recepción, incluyendo los comentarios, y los debates más o menos serenos o de apariencia objetiva. En efecto, la polarización radical de las reacciones de los lectores ante los dos largos poemas dados a conocer entre 1612 y 1617, el *Polifemo* y las *Soledades*, entusiastas muchos y algunos malhumorados o indignados, condicionaron la recepción del poeta y todo lo que tiene que ver con ella, impregnando de sesgo polémico todo cuanto se dijo no solo de estos poemas sino de todo lo escrito por Góngora o por quienes fueron percibidos como de su misma casta o secta estética y literaria. Así sucedió incluso en géneros ajenos a la obra del poeta cordobés, como la predicación, no solo en su tiempo sino hasta mucho más tarde, y en parte hasta hoy. De modo que, por un proceso de progresiva ampliación, decidimos considerar como testimonio de la polémica cualquier escrito que diera testimonio de la lectura de Góngora en su tiempo y cualquier tipo de apreciación crítica acerca de él o de sus presuntos discípulos y admiradores, a los que se solía llamar cultos.

15. <https://obvil.sorbonne-universite.fr/>.

II.1. *El terreno movedizo y las fronteras elásticas de la polémica gongorina*

La importancia que cobra esta recepción polémica dentro del conjunto del sistema literario del siglo XVII confiere a sus fronteras una elasticidad peculiar. Por dar aquí solo un ejemplo, pero significativo, vale la pena recordar que la obra de Lope de Vega, en sus últimos veinte años de vida y tal vez antes, está fuertemente condicionada por los deseos contradictorios de afirmarse como el antagonista de Góngora, su contrario, y el de hacer lo mismo que el poeta cordobés, pero mejor. Dicha actitud generó un puñado de escritos críticos y polémicos que son abiertamente tales y, por otro lado, obras de apariencia neutra (comedias, novelas, poesías líricas) que, además de otros objetivos, contienen alusiones crípticas a este debate, y, en su calidad de imitaciones encubiertas o emulaciones ostentosas, dan testimonio de la recepción de Góngora.¹⁶ Lo interesante en el caso de Lope es que en parte se trata de un debate consigo mismo, una especie de monólogo dramatizado. Lo mismo, hasta cierto punto, sucede con los demás poetas ambiciosos de esos años, de modo evidente con Francisco de Quevedo¹⁷ y Juan de Jáuregui.¹⁸ Los temas y las posiciones expresadas, los enfrentamientos y las alianzas entre poetas y escritores, durante buena parte del siglo, tienen a Góngora como referente, explícito o implícito.

Por ello, nos ha parecido oportuno incluir en el ámbito del proyecto textos como el opúsculo titulado *Anti-Jáuregui del licenciado Luis de la Carrera al reformador de los poetas castellanos*. El texto así llamado, conservado en un manuscrito misceláneo de la Real Biblioteca del Monasterio del Escorial, era respuesta acerada a otro texto polémico, la *Carta del licenciado Claros de la Plaza al maestro Lisarte de la Llana*. Esta *Carta*, que circuló sin autoría a finales de 1624, denunciaba los atentados a la pureza de la lengua castellana supuestamente cometidos por Lope en su epopeya *Jerusalén conquistada*, y permaneció inédita hasta su publicación por Antonio Paz y Meliá en *Sales españolas*, a principios del XIX.¹⁹ Aunque el texto no iba firmado, quienes lo leyeron en su tiempo lo atribuyeron sin vacilar al poeta sevillano Juan de Jáuregui, el mismo que ya se había señalado como crítico agudo, agrio y lenguaraz, unos diez años antes, con el famoso *Antídoto contra la pestilente poesía de las Soledades*. Poco antes de redactar la tal carta satírica contra Lope de Vega, había reavivado la polémica gongorina con un hermoso y bronco poema, el *Orfeo*, y con un sesudo *Discurso poético* que sin nombrar a nadie, se pronunciaba tanto contra Lope como contra Góngora (Jáuregui 2016). Pues bien, el *Anti-Jáuregui*, que responde agresivamente a la *Carta del licenciado Claros de la Plaza*,

16. Véase Blanco (2008) y Gutiérrez Valencia (2021).

17. Cacho Casal (2004); Carreira (2014); Conde Parrado (2021).

18. Véase Blanco (2008) y Gutiérrez Valencia (2021).

19. Jáuregui (1902).

es, como ha demostrado Juan Montero,²⁰ obra del mismo Lope de Vega, que actúa bajo la máscara de un seudónimo, Luis de la Carrera, dirigiendo tiros a Jáuregui pero también defendiéndose a sí mismo en tercera persona y, lo que es más curioso, defendiendo al mismo Góngora. En este escrito (presumiblemente, pues, de Lope), los dos encarnizados adversarios, el popularísimo poeta madrileño, y el andaluz Góngora, son presentados como dos genios poéticos, que vuelan a la misma altura sublime, atacados insolente y estúpidamente por un gozque, un perro ladrador envidioso y de mente estrecha, Jáuregui. Es decir que esta polémica encubierta o enmascarada entre el Fénix de los ingenios y el dotado y neurótico Jáuregui carece de independencia con respecto a la polémica gongorina. Puede considerarse como un apéndice de la misma, no solo porque en ella se enfrentan dos de los principales actores de dicha polémica, Lope y Jáuregui, sino porque sus argumentos apelan a los esgrimidos a favor o en contra de Góngora.

Por estas razones, aceptamos con entusiasmo la propuesta de Juan Montero y Cipriano López Lorenzo de alojar en nuestra plataforma de [l]ólemos la edición del *Anti-Jáuregui* (o más bien reedición, puesto que ya habían editado esta pieza, con un grado menor de rigor y atención, Miguel Artigas y Julián Zarco Cuevas en 1925 y en 2007 Krystof Sliwa). Además de la edición propiamente dicha del texto, incumbe a los editores determinar los metadatos que figurarán en el encabezamiento de la edición en XML-TEI, según una lista compartida por todos los textos publicados por el proyecto: el identificador del texto,²¹ la identificación de los editores científicos, el autor (Lope de Vega, atribuido), el destinatario o dedicatario (Juan de Jáuregui), el título (*Anti-Jáuregui*), la fecha (1624-1625), las fuentes o en este caso la fuente única, el manuscrito del Escorial;²² las ediciones modernas;²³ el texto de Jáuregui al que el *Anti-Jáuregui* responde;²⁴ otros textos

20. Montero (2008). La demostración es reiterada y completada en la introducción del *Anti-Jáuregui* por Cipriano López Lorenzo (véase Vega, 2023).

21. Muy en los primeros tiempos del proyecto e incitado por los ingenieros del OBVIL, Jaime Galbarro García realizó una compleja hoja de cálculo donde aparecían todas las piezas del corpus que nos proponíamos editar (en línea, designados por un identificador) y en columnas una serie de datos: autor, fecha, título, fuentes (en el sentido de testimonios del texto, manuscritos e impresos), ediciones, destinatario o dedicatario, texto al que respondía, referencias a otros textos. La lista fue más adelante progresivamente aumentada por textos nuevos que iban apareciendo. Los identificadores con una configuración fija (en este caso, Blanco, Elvira y Plagnard, 2021, núm. 79. #lopeatribuido_anti-jáuregui), permitían definir los nodos de la red, y la columna encabezada por “Responde a otro texto” las aristas de la red. Esta parte de la hoja de cálculo, una vez convertida por un software adecuado, permite pues generar estas famosas redes. Véase Elvira y Plagnard, 2021 y Pezzini y Ruiz Soto 2017.

22. Real Biblioteca del Monasterio del Escorial, ms. L-1-15, f. 222r-230v, *Anti-Jáuregui del licenciado don Luis de la Carrera al reformador de los poetas castellanos*.

23. Artigas (1925: 587-605), Zarco Cuevas (1925: 272-290), Sliwa (2004: 322-336 y 2007: 112-124).

24. Juan de Jáuregui, *Carta del licenciado Claros de la Plaza al maestro Lisarte de la Llana*, en Jáuregui, 2022.

del corpus al cual hace referencia.²⁵ Todo ello ha sido codificado en TEI y puede servir, como comentamos más abajo, para la exploración informática del corpus, mediante el etiquetado, de manera similar al caso del proyecto PRESOLO que hemos descrito.

En definitiva, al decidir editar las piezas de la polémica gongorina, nos metimos tal vez inadvertidamente en una red de escritos tupida y compleja, estrechamente conectada con textos que, a pesar de tratar temas distintos, demuestran un grado variable de relación con las piezas polémicas más explícitas sobre la poesía de Góngora.

Los textos ya publicados en línea, que dependen de una serie de contingencias —encontrar a los editores dispuestos a trabajar según el patrón que les indicamos, esperar la entrega de sus ediciones, buscar tiempo y mano de obra para tratarlas (corregir, regularizar, codificar)— representan solo una pequeña parte de lo que está en curso o lo que sería deseable hacer. Parece necesario precisar que al decir “nosotros” nos estamos refiriendo a un grupo también plural y abierto; si la iniciativa fue de Mercedes Blanco, la idea tomó forma conversando con varias personas, como Roland Béhar y Begoña López Bueno; entre quienes lanzaron el proyecto, hay que mencionar a Jaime Galbarro y Sara Pezzini; entre quienes generosamente contribuyeron a él, a Héctor Ruiz Soto, Antonio Rojas y Marie-Églantine Lescasse y quien hoy contribuye a hacerlo vivir, es ante todo Aude Plagnard. La lista de textos publicados aparece en la captura de pantalla siguiente (Fig. 2):

↓ Date ↑	↓ Auteur ↑	↓ Titre ↑
1586	Góngora, Luis de	Comedia Venatoria
1588	Pellegrino, Camillo ; Salviati, Lionardo	Lo 'Nfarinato secondo
1610	Góngora, Luis de	Comedia de las firmezas de Isabela
1613	Anónimo	Carta de un amigo de don Luis de Góngora en que da su parecer acerca de las Soledades que le había remitido para que las viese
1613	Góngora, Luis de	Comedia del doctor Carlino
1614	Juan de Espinosa	Carta de Juan de Espinosa a Juan de Arguijo sobre la poesía oscura, con otras sales contra Góngora
1614	Almansa y Mendoza, Andrés de	Advertencias [...] para inteligencia de las «Soledades» de don Luis de Góngora
1614	Fernández de Córdoba, Francisco	Parecer de don Francisco de Córdoba acerca de las Soledades, a instancia de su autor
1616	Ponce, Manuel	Silva a las Soledades de don Luis de Góngora, con anotaciones y declaración, y un discurso en defensa de la novedad y términos de su estilo
1617	Fernández de Córdoba, Francisco	Examen del Antídoto o Apología por las Soledades de don Luis de Góngora contra el autor del Antídoto

25. Blanco, Elvira, Plagnard (2021: núms. 25 [#jauregui_antidoto], 55 [#lope_justa-isidro], 61 [#lope_filomena-jardin], 74. [#lope_relacion-isidro], 77 [#jauregui_discurso-poetico], 78 [#es-trada_papel], 81 [gongora_soneto-orfeo] y 84 [#lope_montalban-orfeo-cancion]).

1617	Suárez de Figueroa, Cristóbal	El pasajero. Advertencias utilísimas a la vida humana
1617	Díaz de Rivas, Pedro	Anotaciones a la Segunda Soledad
1621	Anónimo ; Vega Carpio, Lope de	Epístolas de La Filomena de Lope de Vega
1624	Lope de Vega	Epístolas de Diego de Colmenares y Lope de Vega (en La Circe)
1624	Jáuregui, Juan de	Discurso poético de don Juan de Jáuregui
1624	Faria e Sousa, Manuel de	Noches claras [fragmento]
1624	Anónimo	Contra el Antídoto de Jáuregui y en favor de don Luis de Góngora, por un curioso
1625	Lope de Vega	Anti-Jáuregui
1625	Castillo Solórzano, Alonso de	El culto graduado de Alonso de Castillo Solórzano
1625	Góngora, Luis de	Epistolario
1625	Góngora, Luis de	Poesía
1628	Hernando Horio ; Juan de Pineda	Dos censuras inquisitoriales de las Obras en verso del Homero español (1628)
1628	Paravicino, Fray Hortensio Félix	Vida y escritos de don Luis de Góngora
1629	Suárez de Figueroa, Cristóbal	Pusílipo. Ratos de conversación en los que dura el paseo
1629	Pellicer, José	Vida de don Luis de Góngora
1631	Quevedo, Francisco de	Prólogo a las obras de Fray Luis de León
1633	Anónimo	Escrutinio sobre las impresiones de las obras poéticas de don Luis de Góngora y Argote
1633	González de Salas, Jusepe Antonio	Sección V de la Nueva idea de la tragedia antigua o ilustración última al libro singular de Poética de Aristóteles Stagirita
1633	Paravicino, Fray Hortensio Félix	Vida y escritos de don Luis de Góngora
1634	Cascales, Francisco	Cartas sobre la poesía nueva de don Luis de Góngora
1635	Cuesta, Andrés	Traducción latina de algunos versos del Polifemo
1635	Cuesta, Andrés	Censura a las "Lecciones solemnes" de Pellicer
1635	Angulo y Pulgar, Martín de	Epístolas satisfactorias
1636	Villar, Francisco del	Fragmentos del Compendio Poético
1636	Gómez de Tejada y de los Reyes, Cosme	León prodigioso. Apología moral entretenida y provechosa a las buenas costumbres, trato virtuoso y político. Apólogos XLII y XLIII
1638	Pellicer y Tovar, José	Segundas Lecciones solemnes
1662	Vaca de Alfaro, Enrique	«Don Ludovicus de Góngora y Argote» en Varones ilustres de Córdoba.
1662	Espinosa Medrano, Juan de	Apologético en favor de don Luis de Góngora, príncipe de los poetas líricos de España, contra Manuel de Faria y Sousa, caballero portugués

Fig. 2. Índice del proyecto *Polémos*.²⁶

26. Véase la paleta de proyectos desarrollados en el OBVIL: <https://obvil.sorbonne-universite.fr/projets>, así como el índice del proyecto reproducido en la figura 1: <https://obvil.sorbonne-universite.fr/corpus/gongora/polemos/>.

II.2. Razones hermenéuticas e históricas de *Πólemos*. Más allá de la edición, la construcción de un objeto de conocimiento

Las razones de nuestro interés por editar estos textos eran, como apuntamos, de índole filológica y hermenéutica. La obra de Góngora nos ha llegado envuelta en la fama de haber suscitado, desde el inicio, adhesiones casi fanáticas y rechazos brutales, o pérfidamente cachazudos como el de Cascales.²⁷ El rumor de escándalo se prolonga durante siglos y sus ecos siguen alimentando las proclamas de los ilustrados presuntamente clásicos contra la decadencia de nuestras letras en el siglo barroco. No hay historia de la literatura, trasnochada o puesta al día, que no nos hable de cosas como el cultismo y el culteranismo, supuestos movimientos o tendencias características de una parte sustancial de la literatura española. Ahora bien, estos términos, inexistentes en el siglo XVII, que sí habló de cultos y de culteranos pero no acuñó sus derivados abstractos cultismo y culteranismo, remiten a conceptos o pseudo-conceptos que se interponen entre Góngora y nuestras lecturas de él y de otros escritores de su tiempo. Aunque estos conceptos nos alejan de lo que fue la recepción de Góngora durante su vida y en las décadas inmediatamente posteriores, hay que reconocer que estos no existirían si no hubiera habido una polémica coetánea del poeta que tocó a casi todos los literatos y eruditos de algún renombre, e incluso se convirtió en la máxima referencia crítica y teórica en materia literaria durante la época áurea de nuestras letras. En la segunda década del siglo XXI, cuando lanzamos el proyecto *Πólemos*, la polvareda levantada por detractores y aficionados no se había depositado del todo y seguía orientando, para bien o para mal, nuestra percepción de aquella lejana literatura que para nosotros hace oficio de clásico, en el sentido de que sigue parcialmente viviendo o siendo capaz de revivir.

La convicción que nos movió a constituir un corpus editable con los documentos de la polémica fue que ante esta situación solo cabían dos posibilidades. O bien se hace tabla rasa de todo lo dicho acerca de Góngora en su siglo y en los siguientes, y nos dirigimos a los textos con mirada virgen y cándida, o lo que es lo mismo, desde nuestros gustos y saberes, ignorancias y prejuicios de hoy (algo ajeno a los principios fundamentales de la filología, que no han sido superados ni lo serán nunca, mientras exista esta disciplina). O bien intentábamos historiar la polémica sin prejuicios y simplificando lo menos posible, para invertir precisamente el proceso de comprensión perentoria pero rudimentaria a la cual estábamos abocados los críticos y los lectores en general, quien más o quien menos, por un pseudo-conocimiento apresurado y parcial del fenómeno. Esta segunda vía exigía reunir el mayor número posible de textos y tratarlos con absoluta seriedad, sin precipitación, sin jerarquizarlos a priori, sin pensar que ya sabíamos de antemano lo que tenían que decirnos. Cada uno de esos textos debía ser objeto de una lectura atenta, demorada, que pusiese de relieve y resolviera las muchas dificultades que planteaban. Además

27. Véase la edición de Mercedes Blanco y Margherita Mulas (2018).

de la dificultad propia de cualquier texto del XVII, con su manejo de una lengua rica y de una retórica compleja, hoy fuera de uso, los escritos que nos interesan presentan una dificultad que procede de dos características de distinto signo: por un lado, la erudición humanística, las incesantes referencias a los clásicos griegos y latinos, y por otro el carácter jocosos y maliciosos, chistoso y retorcido, alusivo e ingenioso de todo texto de combate, sobre todo en materia poética.

De ahí que buscáramos ediciones muy cuidadosamente anotadas que resolvieran en lo posible estas dificultades, desplegaran lo que en los textos estaba expresado en forma de insinuación o alusión, ayudaran a seguir la argumentación y aclararan sus presuposiciones, así como tradujeran conceptos ajenos a la cultura de hoy. Por otra parte, se trataba de situar cada texto en una red intertextual de la mayor riqueza posible. Sin encerrarnos en la cuestión debatida en todo el corpus de la validez de Góngora como modelo, afirmada o negada, nos proponíamos examinar la hipótesis de que en estos textos se discuten ciertos conceptos importantes para la definición de la literatura, como su relación con la lengua, y con la evolución histórica del idioma, o el carácter irreductible o idiosincrático que reviste el estilo en los poetas y escritores de primer orden. Pensábamos que la noción de oscuridad poética y el debate sobre si debe ser promovida y condenada, ciertamente importante (Roses 1994), no agota la riqueza conceptual y doctrinal de este puñado de escritos.

Por supuesto, se trataba también de averiguar cuánto fuera posible sobre el contexto en que se compusieron los textos, sobre la cronología, el perfil de cada autor, su relación con Góngora o con los demás polemistas, la cultura o la biblioteca que moviliza, los fines que persigue, su personalidad y sus hábitos como intelectual o como miembro de la República de las letras.

Nuestro objetivo era, pues, de orden estético y de teoría literaria (nueva interrogación de los fundamentos de la literatura, pero no en abstracto sino como parte de una tradición y realidad histórica), y, por otro lado, también de historia social de hechos literarios o paraliterarios, similar en ello a la de PRESOLO, que estudia las “prácticas editoriales y sociabilidad literaria en torno a Lope de Vega” o, de modo más amplio, a las investigaciones llevadas a cabo en el SILEM, *Sujeto e institución literaria en la Edad Moderna* —federación, en realidad, de varios proyectos y equipos importantes—.²⁸

II.3. *Cuestiones de método y medios*

Para acercarse a la consecución de este objetivo, había que crear una plataforma digital que funcionara como un escenario donde todas esas voces se hicieran oír como conjunto o concierto (a veces cacofónico). Las escasas visibilidad y remuneración de este trabajo tan arduo, hicieron necesario movilizar a muchas perso-

28. <https://www.uco.es/redvoces/>

nas, a muchos editores y dedicar un denodado esfuerzo a la relectura, corrección, y uniformización de las ediciones en Word por ellos entregadas y a la codificación en XML-TEI de los archivos que nos remitían.

La empresa se reveló más espinosa de lo que creíamos, pero también, o así lo esperamos, más interesante, aunque solo sea por razones de método. En efecto, nos propusimos poner al mismo nivel los textos y dedicarles el mismo grado de atención, a pesar de su heterogeneidad en múltiples aspectos. Dispare, los textos que editamos lo son en la calidad de su escritura (desde la banalidad hasta la originalidad y los destellos más brillantes), así como también por su extensión, por su grado de notoriedad y accesibilidad. Acerca de esto último, los hay que tienen hasta cuatro o cinco ediciones modernas, como el *Apologético* de Espinosa Medrano, y otros que nadie ha leído, por encontrarse en manuscritos de propiedad privada, como la *Silva* de Manuel Ponce, o incluso en apariencia perdidos, como el inmenso *Antifarriscarcho* de Angulo y Pulgar, que acaba de aparecer.²⁹ También hay otros prácticamente vírgenes por la dificultad misma del documento, como las notas sueltas a Góngora de Martín Vázquez Siruela (1600-1664), escolios a fragmentos gongorinos que se suceden en completo desorden, a través de dos centenares de folios cubiertos por una escritura diminuta. Este borrador inacabado, que se encontró entre los papeles por el erudito granadino después de su muerte, es una mina de ideas y de datos, en la que están trabajando Pedro Conde Parrado y Mercedes Blanco.

Publicar textos tan numerosos y dispares no hubiera sido posible por razones logísticas y económicas en una edición tradicional. En primer lugar, porque era necesario mantener abierto el proyecto durante bastantes años —ya llevamos casi diez, y hay tela cortada para otros cinco más como mínimo—, no solo para añadir textos y corregir los textos existentes a la luz de las nuevas publicaciones, sino también para introducir nuevas funcionalidades que permitan enriquecer el trabajo editorial. Esa conciencia de que nada se acaba nunca de verdad en cuestiones de investigación y de trabajo, en especial en proyectos editoriales, tiene grandes ventajas, porque permite darse objetivos ambiciosos. Como escribe Susanna Allés-Torrent, la edición se convierte “en un proceso y no tanto en un producto”.³⁰ También tiene, por supuesto, inconvenientes, y puede ser tachada de irrealista.

Por otra parte, nuestra intención fue hacer posible una lectura conectada de los textos, para lo cual impusimos una serie de normas homogéneas que hicieran posible circular de varias maneras dentro del corpus. El sentido del proyecto consistía, en efecto, en volver accesibles y visibles el mayor número de conexiones, tanto de contigüidad como de similitud, entre los diferentes textos. Por

29. Este último manuscrito se creía perdido y acaba de aparecer a raíz de una sucesión en Córdoba; lo ha comprado hace pocos meses la Biblioteca de la Universidad de Sevilla por mediación de Begoña López Bueno.

30. Allés-Torrent (2020), de hecho, cita a Patrick Sahle (2016: 29), quien habla elegantemente de “fluid publication”.

ello, nos esforzamos en crear toda clase de paralelismos y pasarelas entre ellos. Solo citaré tres aspectos de esta homogeneización.

A. Las introducciones

Las introducciones se formalizaron siguiendo un patrón único, como puede observarse en estos dos ejemplos, lo cual facilita una lectura transversal y la comparación de las características de los textos gracias a los distintos epígrafes (*cf.* Fig. 3):



Fig. 3. Estructuras de las introducciones: ejemplos del *Examen del Antídoto* (ed. Matteo Mancinelli)³¹ y del *Anti-Jáuregui* (ed. Juan Montero et Cipriano López Lorenzo).³²

31. https://obvil.sorbonne-universite.fr/corpus/gongora/1617_examen.

32. https://obvil.sorbonne-universite.fr/corpus/gongora/1624_anti-jauregui.

B. Ecdótica y modernización

No hemos aplicado métodos de la ecdótica renovados por la herramienta digital; en este aspecto hemos dejado plena libertad a los editores, por lo general no formados en esta herramienta o no especialmente interesados por ella. Muchas veces el texto se ha transmitido por un solo testimonio (manuscrito o impreso). Cuando son varios, se han cotejado para establecer un texto crítico. El proceso ha sido ejemplarmente riguroso en el caso de la edición del *Examen del Antídoto* de Matteo Mancinelli, con ocho testimonios manuscritos, lo que ha dado lugar a proponer un *stemma* y a aplicarlo al texto crítico.

Lo que ha sido en cambio condicionado por el medio digital han sido las normas de modernización, bastante drásticas, según el principio de imponer una homogeneidad que permita las búsquedas y que de otro modo sería imposible o mucho más complicada.³³

Esta modernización consiste en la regularización, de acuerdo con la norma actual, de todas las fluctuaciones gráficas: s/ss; ç/c/z; x/j/g; u/v; v/b; r/rr; qu/cu; m/n delante de p/b; la supresión de los grupos consonánticos cultos ph, th y ch, y la modernización de los grupos cultos con pertinencia fonológica (pt, ct, gn, bst, nct, sc); la eliminación del vocalismo átono, la separación de las aglutinaciones (desta > de esta; esotra > esa otra; estotros > estos otros; della > de ella; dél > de él); el desarrollo de las abreviaturas sin indicarlo (V. M. > vuestra merced); la restauración de la h-etimológica, y la modernización de la puntuación y acentuación. Se conserva, en cambio, la estructura de vocablos como *aquesto* o *aquese* que no tienen equivalente exacto. También hay que modernizar por completo los nombres propios, desechando en ellos todo tipo de grafías antiguas. Sin embargo, para la edición de los versos castellanos (de Luis de Góngora o de otros autores, ya sean originales o traducidos), aplicamos una modernización moderada con el fin de respetar la configuración rítmica del verso.

C. Anotación

La perspectiva de una publicación digital de los textos nos permitió, por otra parte, idear un sistema de anotación con varios niveles, en función de la calidad de la información anotada. Concretamente, introdujimos tres niveles de notas: las “de autor”, las de “edición crítica” y las de “edición”.

Las notas de autor tienen como función presentar las anotaciones que los autores (o en algunos casos los editores) introdujeron en sus propios textos, gene-

33. No deja de haber una tensión (que puede ser contradictoria) entre el carácter plural, fluido y dinámico del texto que induce el tratamiento digital (en el que, según dicen ingeniosamente algunos, Heráclito y su materialismo han desplazado el idealismo platónico) y la necesaria estabilidad e incluso rigidez que exigen los sistemas automáticos de búsqueda, la estilometría, la minería de datos, y todos los acercamientos cuantitativos a textos o corpus. Véase Pierazzo (2016: 56). Las matemáticas siempre han sido platónicas.

ralmente con referencias eruditas. Decidimos presentar estas notas por eskeumorfismo, es decir imitando, visualmente, los marginalia del texto editado (cf. Fig. 4):

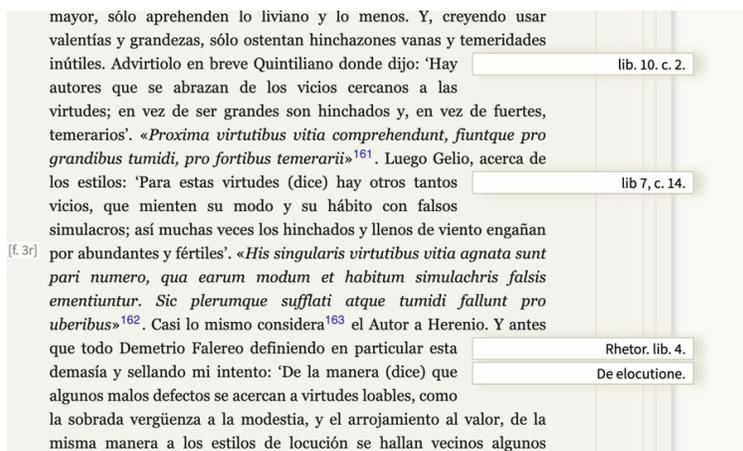


Fig. 4. Notas de autor: el ejemplo del *Discurso poético* de Juan de Jáuregui (2016) en la edición de Mercedes Blanco³⁴

Las notas de edición crítica dan un aparato positivo de variantes o informan sobre algún accidente del testimonio y alguna decisión del editor para compen-sarlo. Las correspondientes llamadas de notas son números romanos.

Las notas de edición, las más abundantes y corrientes, son notas explicativas del editor, y cumplen con varios propósitos: aclarar dificultades léxicas o sintácticas, desplegar las referencias eruditas, dar traducciones de los textos latinos o en otras lenguas extranjeras, suplir el carácter elíptico o alusivo del texto, proporcionar datos contextuales imprescindibles para la comprensión de determinado fragmento.

III. Características de la edición XML-TEI

Las ediciones que proponemos en el marco del proyecto Πólemos no solo se ajustan a la gramática de la Text Encoding Initiative. También, por insertarse en el Labex OBVIL, arriba descrito, siguen una serie de pautas y herramientas crea-das en el marco de dicho Labex. Empezaremos por describir este marco colecti-vo, en la medida en que determina algunos aspectos de la estructura de las ediciones, antes de presentar las adaptaciones que hicimos para dotar a Πólemos de unas herramientas específicas.³⁵

34. https://obvil.sorbonne-universite.fr/corpus/gongora/1624_discurso-poetico.

35. Sobre estas cuestiones, véase la luminosa síntesis de Susanna Allés-Torrent (2015).

III.1. *Los límites de la conversión automática*

En la primera fase del proyecto, nuestro equipo trabajaba en ficheros .docx y la conversión al TEI corría a cargo de los informáticos del OBVIL. Frédéric Glorieux había creado dos tipos de programas para semi-automatizar la conversión del formato editorial (.doc) a los distintos formatos de publicación que ofrecía el OBVIL (tei, html, epub, kindle, txt e iramuteq):

1. Odette,³⁶ un programa que convierte los documentos del formato ODT al formato TEI;
2. Una librería de programas libres XSLT/PHP titulada *Teinte suite* destinada al conjunto del proyecto y que aseguraba la conversión al html y a los demás formatos.

Esta opción metodológica y estas herramientas estaban pensadas para dissociar las competencias de edición científica y las de edición digital. Traduzco a continuación la presentación que Frédéric Glorieux hizo de esta herramienta:

Odette le garantiza al LABEX OBVIL el control de una de las etapas decisivas en su cadena de producción de textos. En un contexto universitario, existe la posibilidad de movilizar a muchos colaboradores bien formados (estudiantes, doctorandos, investigadores) de forma puntual. Enseñarles el lenguaje XML-TEI representa una formación a largo plazo, para la que suelen estar poco motivados porque es una competencia que raramente podrán volver a movilizar. El OBVIL invirtió entonces en un uso avanzado del tratamiento de textos (creación y aplicación de estilos) [...] [para] que un colaborador se vuelva operativo de modo más rápido, en base a una competencia que le resultará necesaria en un futuro, de forma que perfecciona sus prácticas. [...] El propósito es hacer entender que la estructuración informática de un texto no es un artesanado subalterno, reducible a un trabajo de secretaría editorial cosmética, sino que constituye una responsabilidad intelectual de la que debe hacerse cargo un autor o un editor.³⁷

Se trataba pues de optimizar la operatividad de los editores científicos, aunque reduciendo su intervención al ámbito del tratamiento de texto, y apostando por la conversión automática del texto del .doc al .xml

Los programas de conversión estaban destinados a satisfacer las necesidades de ediciones aplicadas a corpus muy distintos entre sí,³⁸ con la idea de simplificar el proceso de conversión a la escala del Labex. Esta práctica, que funcionaba hasta cierto punto a dicha escala, pronto demostró sus límites en la medida en que tuvimos que solicitar una serie de funcionalidades específicas que implicaban diseñar un etiquetado propio para nuestro proyecto.

36. <https://obvil.sorbonne-universite.fr/developpements/odette>.

37. Traducimos a Glorieux (2018).

38. Véase la paleta de proyectos desarrollados en el OBVIL: <https://obvil.sorbonne-universite.fr/projets>.

Frente a esta situación, decidimos asumir un trabajo de edición y de etiquetado XML-TEI más fino que el que nos permitían estos programas de conversión y seguir colaborando con los ingenieros para que adaptaran las hojas .xsl a nuestro proyecto, tanto para la publicación en HTML como para la extracción de datos. Dominar el TEI supuso un salto cualitativo para nuestras ediciones, aunque nos vemos todavía limitados por el hecho de no ser autónomos en el manejo de la escritura en .xsl.

III.2. Estructura general de nuestras ediciones

Nuestras ediciones siguen la estructura clásica de los documentos TEI: un encabezado, o <teiheader>, que recoge los metadatos, seguido por el cuerpo (o <body>) de la edición. Este último se subdivide a su vez en dos partes principales, la primera para la introducción —con su arborescencia de apartados y subapartados que reflejan la estructura común a todas nuestras ediciones, tal y como la describimos más arriba— y la segunda para el texto de la edición —con divisiones variables, en función de la propia estructura del texto—. El índice de las *Epístolas satisfactorias* de Martín de Angulo y Pulgar, visto en edición TEI, proporciona un buen ejemplo de esta estructura:

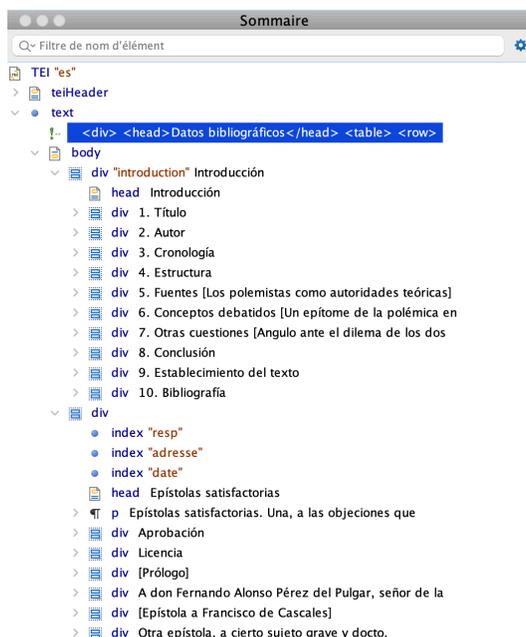


Fig. 5. Índice en TEI de la edición por Juan Manuel Daza las *Epístolas satisfactorias* de Martín de Angulo y Pulgar (2018).³⁹

39. https://github.com/gongoradigital/polemos/blob/gh-pages/1635_epistolas.xml.

El TEI-header de nuestras ediciones consta de los siguientes elementos: <fileDesc>, <encodingDesc> y <profileDesc>.

<fileDesc> describe las características del texto editado (título o <titleStmt> y fuentes en las que se basa la edición, descritas en el <sourceDesc>) y de la edición (<editionStmt> y <publicationStmt>).

<encodingDesc> recapitula a su vez los tres niveles de notas que identificamos anteriormente:

L'édition comporte trois niveaux de notes. note[@place="bottom"] : notes d'éditeur (décimal par défaut) note[@place="margin"] : notes marginales de l'auteur (pas de numérotation) note[@type="app" rend="1"] : notes d'apparat (I = romain).⁴⁰

En la edición en Word proporcionada por los editores, las notas a pie de página iban precedidas de una marca que indicaba su categoría ([NDA], [NDEC], [NDE]) diferencias que son traducidas automáticamente al lenguaje de marcado XML-TEI y su procesamiento y presentación resultan mucho más ágiles y cómodos que en los métodos tradicionales.

Finalmente, <profileDesc> recapitula la fecha de creación del documento y su idioma.

TEI	teiHeader	fileDesc	editionStmt
5	TEI	xmlns="es"	xmlns="http://www.tei-c.org/ns/1.0" n="1613_carta-espinoza">
6	<teiHeader>		
7	<fileDesc>		
8	<titleStmt>		
9	<title>	Carta de Juan de Espinosa a Juan de Arguijo sobre la poesía oscura, con otras sales	
10	</title>		
11	<author>	Juan de Espinosa</author>	
12	<editor>	Héctor Ruiz Soto</editor>	
13	</titleStmt>		
14	<editionStmt>		
15	<edition>	OBVIL</edition>	
16	<resp>	Mercedes Blanco</name>	
17	</resp>		
18	<resp>	Muriel Elvira</name>	
19	</resp>		
20	<resp>	Aude Plagnard</name>	
21	</resp>		
22	<resp>	Héctor Ruiz Soto</name>	
23	</resp>		
24	<resp>	estylage et édition TEI</resp>	
25	</resp>		
26	<publication>		
27	<publisher>	Université Paris-Sorbonne, LABEX OBVIL</publisher>	
28	<date>	2021</date>	
29	<availability>	restricted</availability>	
30	<license>	https://creativecommons.org/licenses/by-nc-nd/3.0/fr/</license>	
31	</license>		
32	<sourceDesc>		
33	<source>		
34	<title>	Libro de varios tratados de prociiosidad y erudición, de diferentes	
35	</title>		
36	<author>	Héctor Ruiz Soto, Universidad de Sevilla, ref	
37	</author>		
38	<target>	https://archive.org/details/IR00158/page/n298/mode/1up	
39	</target>		
40	<sourceDesc>		
41	<encodingDesc>		
42	<note>	note[@place="bottom"] : notes d'éditeur (décimal	
43	</note>		
44	<note>	note[@place="margin"] : notes marginales de l'auteur (pas de numérotation)	
45	</note>		
46	<note>	note[@type="app" rend="1"] : notes d'apparat (I = romain)	
47	</note>		
48	<profileDesc>		
49	<creation>	date notBefore="1613" notAfter="1614">	
50	</creation>		
51	<language>	es</language>	
52	</language>		
53	</profileDesc>		
54	</teiHeader>		

Fig. 6. TEI-header de la *Carta de Juan de Espinosa a Juan de Arguijo* (2021), editada por Héctor Ruiz.⁴¹

⁴⁰. Esta mención se encuentra en el <encodingDesc> de todas nuestras ediciones.

⁴¹. https://github.com/gongoradigital/polemos/blob/gh-pages/1662_apologetico.xml.

III.3. *Editar la obra de Góngora*

Nuestra idea, cuyo desarrollo presenta todavía bastantes deficiencias, fue utilizar el corpus de textos polémicos y comentarios como una caja de resonancia para el texto mismo de Góngora, algo que permitiera pasar de sus versos a las prosas de sus comentaristas, detractores y defensores. Para ello editamos la obra completa de Góngora, codificando en TEI una versión de la edición de esta obra por Antonio Carreira, revisada por el editor: o sea las poesías de atribución segura o muy probable, el teatro y el epistolario. En esta edición se pueden realizar búsquedas en la obra de Góngora con mayor comodidad y exhaustividad que cualquier otro método disponible, puesto que pueden buscarse grupos de palabras, palabras cercanas, o cualquier otro tipo de cadena de signos. Por lo demás, aparecen al margen de la edición de cada poema llamadas de nota a lugares en que estos versos son comentados. Por ejemplo, pueden leerse, al margen del texto de Góngora las anotaciones de Díaz de Rivas a modo de llamadas de notas marginales. En el marco de este trabajo, nos centraremos exclusivamente en el etiquetado de la poesía, que es la parte del corpus que más conexiones presenta con la polémica.

Para la poesía, publicada en 2016, compusimos un documento formado de una división (<div>) por poema y numeramos con un identificador único (<xml:id>) cada poema y cada verso. Los poemas fueron numerados conforme al modelo de la edición (de 1 a 418 para los poemas de atribución segura, y 419 a 477 para aquellos de atribución probable) e introdujimos subdivisiones para distinguir las subpartes formadas por la dedicatoria y el texto de la *Soledad primera* y el texto de la *Soledad segunda* (respectivamente, xml:id="poem264A", xml:id="poem264B" y xml:id="poem264C"). En efecto, estas tres piezas, reunidas bajo el mismo número 264 en la edición de referencia, quedan generalmente identificadas por separado en los discursos de los polemistas. A su vez, los versos se numeraron (xml:id="l264A.1"), permitiendo las referencias concretas a cada uno de ellos. Finalmente, identificamos, aunque sin numerarlas, las estrofas que componen cada poema (cf. Fig. 7). Este último etiquetado podría servir de base, en un futuro, para la identificación automática de las secuencias métricas, tarea en la que no nos hemos lanzado todavía.

19636	<div n="264A" type="poem" xml:id="poem264A">
19637	<index>
19638	<term type="head"><num>264a. </num>Dedicatoria Soledades</term>
19639	</index>
19640	<head type="sur"><num>264a</num></head>
19641	<head type="main" rend="center">Soledades</head>
19642	<head>Al duque de Béjar</head>
19643	
19644	<lg>
19645	<l n="1" xml:id="l264A.1">Pasos de un peregrino son, errante,</l>
19646	<l n="2" xml:id="l264A.2">cuantos me dictó versos dulce musa,</l>
19647	<l n="3" xml:id="l264A.3"> en soledad confusa</l>
19648	<l n="4" xml:id="l264A.4">perdidos unos, otros inspirados.</l>

Fig. 7. Luis de Góngora, Dedicatoria de la *Soledad primera* (2016) – versión TEI.⁴²

42. https://github.com/gongoradigital/gongoraobra/blob/gh-pages/gongora_obra-poetica.xml.

IV. Hacia una lectura distante del corpus

Las ediciones XML-TEI así establecidas nos permitieron desarrollar herramientas para explorar el corpus en su conjunto. Para guiar esta exploración, nos planteamos tres preguntas:

- ¿Qué textos de Góngora fueron objeto de debate en el marco de la polémica?
- ¿Qué obras y autoridades (antiguas o modernas), fueron convocadas para argumentar en estos debates?
- ¿A qué autores contemporáneos, también involucrados en la polémica, se responde en cada texto? ¿A quién se apoya y defiende o contra quién se argumenta?

Para responder estas preguntas, desarrollamos varias herramientas apoyadas en la misma cadena de codificación y recolección de los datos, según cuatro etapas sucesivas:

1. Anotación filológica
2. Etiquetado en TEI
3. Escenario de transformación para la recolección de los datos etiquetados
4. Puesta en base de los datos recolectados para la constitución de tablas relacionales y gráficos.

IV.1. De la anotación al etiquetado

La anotación filológica de las ediciones permite descifrar las alusiones, referencias y citas contenidas en el discurso del polemista, que nuestros editores anotan, y comentan en el apartado “Fuentes” de la introducción. Si bien las indicaciones bibliográficas representan un primer grado de homogeneización de la información, el etiquetado nos permite sistematizar la identificación de estos datos y recopilarlos de forma automática. Ya lo hemos realizado para dos casos concretos: los versos de Góngora citados en la polémica y las autoridades alegadas para respaldar la argumentación.

En cuanto a los primeros, identificamos, en la anotación, las referencias a poemas o versos de Góngora que son objeto de un debate mediante un identificador: las mayúsculas “OC” seguidas por el número de poema y de verso tal y como aparecen numerados en la *Obra poética completa* de Góngora editada por Carreira que nos sirvió de edición de referencia. Esta información fue luego etiquetada parcialmente en las ediciones TEI, limitándonos a los versos citados textualmente. En este caso concreto, incluimos una etiqueta <ref> dentro de la etiqueta de cita (<quote>), con la doble finalidad de marcar la información para

su futura extracción y crear un enlace que permite navegar de la edición del texto polémico a la de la obra poética de Góngora. Este tipo de etiquetado es especialmente valioso para los textos que citan mucho a Góngora, máxime cuando su propia estructura depende de estas citas, como es el caso de los comentarios, generalmente estructurados poema por poema, estrofa por estrofa y verso por verso. Damos a continuación el ejemplo del etiquetado de las *Anotaciones a la Segunda Soledad* de Pedro Díaz de Rivas, editado por Melchora Romanos y Patricia Festini, en TEI (Fig. 8) y en HTML (Fig. 9):

```

3453 <div>
3454 <head>1.<note place="bottom">En esta primera nota, en el manuscrito, aparece el número uno
3455 después del verso. Como se trata de un caso único, lo transcribimos siguiendo el modelo
3456 del resto de la anotación.</note>
3457 <quote rend="center"><l<ref xml:id="ref1001"
3458 target="http://obvil.sorbonne-universite.fr/corpus/gongora/gongora_obra-poetica/poem264C#1264C.1"
3459 >Éntrase el mar por un arroyo breve</ref></l><note place="bottom">
3460 OC264C.1</note><note place="bottom"> Díaz de Rivas (en adelante, DR) inicia su
3461 comentario a la <hi rend="i">Soledad Segunda</hi>, poniendo de relieve el uso de
3462 la descripción de la ría y ponderando la elaboración poética del fragmento; y, de
3463 acuerdo con el criterio defensivo con que opera, se apoya en autoridades clásicas
3464 y modernas para sostener su valoración positiva de
3465 Góngora.</note></quote></head>
3466 <p>Convenientísimo principio es una descripción para entrar en la narración de cualquiera
3467 obra poética, y así por ella comenzaron gravísimos escritores. El nuestro describe aquí
3468 un arroyo que entraba en el mar, desatando el de su fecundidad, elegancia y erudición en
3469 pintarlo, y aunque parece algo larga la descripción, defiéndase con gravísimos poetas que
3470 a veces se dilatan sólo por desplegar las velas de su elocuencia, ostentando la bazarria y
3471 fecundidad de su ingenio como lo hicieron <persName type="authority">Lucano</persName>,
3472 <persName type="authority">Virgilio</persName>, <persName type="authority">
3473 <persName type="authority">Claudio</persName>, el <persName type="authority">Tasso</persName>, etc.</p>
3474 </div>

```

Fig. 8. Pedro Díaz de Rivas (2017) anotando la *Soledad segunda*, editada por Melchora Romanos y Patricia Festini – versión TEI.⁴³

1.²⁴⁶

Éntrase el mar por un arroyo breve

247 248

Convenientísimo principio es una descripción para entrar en la narración de cualquiera obra poética, y así por ella comenzaron gravísimos escritores. El nuestro describe aquí un arroyo que entraba en el mar, desatando el de su fecundidad, elegancia y erudición en pintarlo, y aunque parece algo larga la descripción, defiéndase con gravísimos poetas que a veces se dilatan sólo por desplegar las velas de su elocuencia, ostentando la bazarria y fecundidad de su ingenio como lo hicieron Lucano, Virgilio, Claudio, el Tasso, etc.

Fig. 9 Pedro Díaz de Rivas (2017) anotando la *Soledad segunda*, editada por Melchora Romanos y Patricia Festini – versión HTML.⁴⁴

43. https://github.com/gongoradigital/polemos/blob/gh-pages/1617_soledad-segunda-diaz.xml
 44. https://obvil.sorbonne-universite.fr/corpus/gongora/1617_soledad-segunda-diaz

En el ejemplo arriba ilustrado, pasamos de la mención “OC264C.1”, que identifica el primer verso de la *Soledad segunda*, al enlace https://obvil.sorbonne-universite.fr/corpus/gongora/gongora_obra-poetica/poem264C#l264C.1, que remite directamente al mismo en nuestra edición.

También decidimos sistematizar el etiquetado en lo que toca a los autores referidos en los textos. Estos pueden ser de dos tipos: por un lado, autoridades antiguas, medievales o contemporáneas, traídas a colación para apoyar un argumento; o bien, y por otro lado, contemporáneos de la República de las Letras involucrados en la polémica y citados en los textos como aliados o como antagonistas. Las autoridades forman una parte esencial de la anotación filológica. Son objeto de una detenida atención en las notas de edición (NDE) que identifican las citas o alusiones, en el apartado “Fuentes” de la introducción que las analiza, y también en lo que denominamos la “biblioteca hipotética del polemista”. Llamamos así el primer apartado de la bibliografía, donde aparecen las “Obras citadas o consultadas por el polemista”. En cuanto a los contemporáneos, más o menos involucrados en la polémica, sus menciones son menos frecuentes, aunque altamente significativas.

Para identificar a uno y otro tipo de autores, utilizamos la etiqueta <persName>, completada por una serie de atributos:

- @authority o @polemist, para especificar si se trata de una autoridad o de un personaje implicado en la polémica;
- @key, para identificar de forma unívoca al autor citado, que en el texto puede encontrarse designado por una multitud de nombres, pronombres o perífrasis;
- en el caso de las autoridades, añadimos el atributo @precision, que nos permitió establecer una casuística para distinguir las autoridades citadas de aquellas simplemente mencionadas o aludidas, entre las citadas, las que lo son directa o indirectamente, es decir, de primera mano, o mediante poliantes u obras de divulgación; son muchas las citas de este último tipo, como lo vemos en el caso de Manuel de Faria e Sousa que, en las *Noches claras*, a partir de un puñado de libros de tipo enciclopédico despliega en boca de sus personajes una erudición de pura fachada.⁴⁵
- de nuevo, en el caso de las autoridades citadas textualmente, añadimos los atributos @unit=”quotes” y @quantity para ponderar estas citas.

Con este etiquetado, obtuvimos una identificación unívoca, homogénea y sistemática de los fenómenos aquí descritos, que traduce el detallado análisis de los filólogos editores al lenguaje de las etiquetas XML.

45. Faria e Sousa (2019).

IV. 2. *Escenario de transformación y análisis de redes*

Este etiquetado nos permitió cartografiar, en base a un análisis de redes, estos fenómenos y analizarlos de manera cuantitativa y cualitativa.

Construimos ciertas herramientas artesanales para la minería de datos y la lectura distante de este amplio corpus, que figuran en la sección “gongotools” de nuestro github:⁴⁶ las hojas .xsl “polemos2nodes” y “polemos2edges” forman un escenario de transformación que permite extraer los metadatos relativos a los poemas de Góngora citados, a las autoridades y a los polemistas, con el fin de circular en el corpus mediante el análisis de redes.⁴⁷ Los xsl compilan automáticamente dos tablas, “graph-to-gongora” y “graph-to-authorities”, adaptadas para el programa de análisis de redes Gephi. A partir de estas tablas y aplicando Gephi, cartografiamos por un lado los poemas de Góngora citados por los textos de la polémica, y por otro las autoridades citadas. A partir de ahí, ha sido posible construir una red de citas para situar los textos unos respecto a otros, en función de su grado de centralidad, como ya hicimos en un artículo del 2018.⁴⁸

El gráfico relacional de textos polémicos y autoridades mencionadas permite distinguir entre las autoridades más citadas, que pertenecen a la cultura común —Virgilio, Horacio, Homero, Marcial, Quintiliano, Cicerón o Garcilaso—, y las que, por el contrario, manifiestan una erudición más recóndita y especializada. Como indicábamos en otro lugar, constatamos que Francisco del Villar (*Compendio poético*) y Lope de Vega (en su correspondencia con Colmenares y en la *Filomena*) comparten un mismo plantel de autoridades humanistas: Justo Lipsio, Cipriano Suárez y Francisco Sánchez de la Brozas.⁴⁹

Evidentemente, estas herramientas son solo una pequeña muestra de las posibilidades de exploración de este riquísimo corpus digital. Aunque su estado actual de desarrollo es limitado, nos permiten medir la utilidad hermenéutica de *navegar* entre estos textos mediante distintas vías.

46. Nuestro repertorio Github, “gongoradigital” (<https://github.com/gongoradigital>) recoge todo el material actualmente disponible del proyecto [I]ólemos. Consta de once repositorios. Tres de ellos contienen los textos publicados en la página del Labex OBVIL: “polemos” reúne las ediciones de la polémica; “gongoraobra”, el conjunto de la polémica gongorina; y “salviati-tasso” alberga un proyecto afín al nuestro, la edición de uno de los testimonios de la polémica tassiana por Chris Geekie. Otros cuatro contienen herramientas de trabajo: “img” almacena las imágenes publicadas por el proyecto; “gongotools” reúne las herramientas de exploración distante del corpus; mientras que “polemos-naked” propone una versión sin etiquetas del corpus, para su exploración por programas de textometría; “teinte” contiene los programas de transformación que usamos en el proyecto. Los otros tres son dedicadas a la formación interna o a la comunicación del proyecto.

47. Otra herramienta de Gongotools: una serie de xsl que prepara el corpus para su exploración en programas de textometría (iramuteq o txm). Se trataba, en este caso, de suprimir toda la información editorial. El tema ha sido explorado por Lescasse (2018).

48. Plagnard y Ruiz Soto (2018).

49. Elvira y Plagnard (2021: 53 y ss).

IV. 3. *Navegar en el corpus*

Para terminar, nos centraremos en este último aspecto, pensando tanto en la perspectiva del editor digital como del lector de la edición frente a su pantalla. Construimos dos herramientas sencillas para este propósito: por un lado, la retrolectura de la polémica a partir de la poesía de Góngora; por otro, una reflexión sobre la intertextualidad en la polémica.

El “graph-to-gongora” descrito arriba (que recapitula las citas del corpus gongorino en los textos polémicos) nos permite reconstituir la red de los poemas de Góngora citados por los polemistas a lo largo del tiempo, es decir, leer a Góngora en función del desarrollo de la polémica. Procediendo al revés, empezamos a construir una versión de la obra de Góngora —organizada cronológicamente, como el manuscrito Chacón y la *Obra* completa editada por Carreira— que sirviera de punto de partida para la lectura de la polémica. Al margen de cada poema o de cada verso, introdujimos enlaces que remiten a los lugares de textos polémicos que citan o comentan dicho fragmento. La herramienta es de las más sencillas y tan solo estamos iniciando este proceso —véase Fig. 10—, que merecería beneficiarse de la edición de los comentarios, piezas inseparables de la polémica.

The image shows a digital interface for a poem. On the left is a vertical list of line numbers from 1590 to 1626. The main area contains the text of the poem, with line numbers 5, 10, 15, 20, 25, and 30 on the left margin. On the right, there are three blue-bordered boxes containing links to polemical texts: 'Carta de un amigo de Góngora', 'Parecer de Francisco de Córdoba', and 'El culto graduado, Alonso de Castillo Solórzano, 1625'. The poem text is as follows:

264B [Soledad] Primera

Era del año la estación florida
 en que el mentido robador de Europa
 (media luna las armas de su frente,
 y el Sol todo los rayos de su pelo),
 5 luciente honor del cielo,
 en campos de zafiro paze estrellas,
 cuando el que ministrar podía la copa
 a Júpiter mejor que el garzón de Ida,
 10 náufrago y desdeñado, sobre ausente,
 lagrimosas de amor dulces querellas
 da al mar; que condolido,
 fue a las ondas, fue al viento
 el mísero gemido
 segundo de Arión dulce instrumento.
 15 Del siempre en la montaña opuesto pino
 al enemigo noto
 piadoso miembro roto,
 breve tabla, delfín no fue pequeño
 al incon siderado peregrino
 20 que a una Libia de ondas su camino
 fió, y su vida a un leño.
 Del océano, pues, antes sorbido,
 y luego vomitado
 no lejos de un escollo coronado
 25 de secos juncos, de calientes plumas,
 alga todo y espumas,
 halló hospitalidad donde halló nido
 de Júpiter el ave.
 Besa la arena, y de la rota nave
 30 aquella parte poca
 que lo expuso en la playa dio a la roca,
 que aun se dejan las peñas
 lisonjear de agradecidas señas.
 Desnudo el joven, cuanto ya el vestido

Fig. 10. Luis de Góngora, “Soledad primera”, anotada por los polemistas, 2016.⁵⁰

50. https://obvil.sorbonne-universite.fr/corpus/gongora/gongora_obra-poetica#poem264B.

Esta herramienta ofrece, sin embargo, un observatorio útil sobre los debates, que sistematiza el que ya encontrábamos —aunque de forma muy fragmentaria— en algunas ediciones del poeta, por ejemplo en la de las *Soledades* por Robert Jammes, que remite con frecuencia a distintos lugares de los comentarios y de las piezas polémicas.

Para cerrar esta reflexión, queremos explorar —aunque sea, de momento, en términos bastante especulativos— la modelización de la intertextualidad en la polémica gongorina. Tal es, en efecto, la cuestión central para la exploración de los debates: ¿quién responde a quién, qué texto a qué texto, en base a qué argumentos y acerca de qué poemas?

Evidentemente, el etiquetado de los polemistas mencionados al hilo de los textos ya proporciona algunas pistas. Este método tiene interés limitado, sin embargo, para abarcar el fenómeno, porque reduce la intertextualidad a los ataques *ad hominem*, sin tomar en cuenta el texto concreto al que responden. Si bien es cierto que tales ataques, muy explícitos y deliberadamente violentos, existieron y constituyen una de las modalidades de la polémica —pensemos en el *Anti-Jáuregui* atribuido a Lope de Vega o en el *Anti-faristarco* de Martín de Angulo y Pulgar—, el carácter anónimo de muchos textos, por razones de censura o de estrategia, complica el proceso.

Por otra parte, las relaciones de intertextualidad son muchas veces más difíciles de describir e implican recurrir a matices más finos. Tomemos el ejemplo de la llamada “Carta echadiza”: esta carta anónima “defiende a Lope de presuntos ataques de Góngora, presentando al primero como inocente y al segundo como malicioso e injusto. Se suele atribuir al mismo Lope []”.⁵¹

Ahora bien, dicha misiva está conectada con una amplia red de textos. En ella, Lope (suponiendo que sea él el autor) *reacciona* a un Góngora que se quejaba, en una carta perdida a Almansa difundida por este en la corte,⁵² de las acusaciones lanzadas por el Fénix en el soneto “Tus bellos ya me matan ojos”.⁵³ En respuesta, Lope sostiene que el soneto no apuntaba a Góngora sino a sus malos imitadores y alega como prueba de su admiración por don Luis otro soneto, “Canta, cisne andaluz, que el verde coro”.⁵⁴ Este intercambio se inscribe en la continuidad de la anterior correspondencia entre el círculo de Lope y el de Góngora⁵⁵ y describe con sorna la conmoción que el “Antídoto” de Jáuregui⁵⁶ provocó entre los defensores de Góngora, burlándose una vez más de las “Advertencias” de Andrés de Almansa y Mendoza.⁵⁷ Por la fecha,

51. *Ibid.*, núm. 34, 575.

52. *Ibid.*, núm. 33, 574.

53. *Ibid.*, núm. 32, 574.

54. *Ibid.*, núm. 30, 573.

55. *Ibid.*, núms. 9-12, 564-566.

56. *Ibid.*, núm. 25, 571.

57. Blanco, Elvira y Plagnard (2021: núm. 8, 564).

puede que Lope ya tenga en mente las defensas de Amaya⁵⁸ y del Abad de Rute.⁵⁹ Como vemos, son muy variados los tipos de relaciones establecidos entre la “Carta echadiza” y esta amplia red de textos. Este acercamiento sutil a la intertextualidad se presta sin duda más a una lectura cercana, cualitativa, que a una lectura cuantitativa.

Con todo, Muriel Elvira y Aude Plagnard han generado un primer intento de modelizar las relaciones intertextuales a escala de toda la polémica —por lo menos, tal y como se conoce en la actualidad—, seleccionando un único fenómeno intertextual: el de “los textos que suscitaron respuestas críticas directas o merecieron apoyo y defensa por parte de sus contemporáneos”,⁶⁰ es decir, aquellos que desempeñaron un papel motor en la polémica. De este modo, conseguimos identificar redes de textos que dialogan entre sí de forma inequívoca y constituyen etapas de mayor o menor duración en el desarrollo de la polémica, tal y como podemos observar en la siguiente figura (11): en azul, la polémica epistolar que constituye el episodio más temprano del que tenemos noticia; en rojo, los debates en torno a los textos de Jáuregui, algo más tardíos; en celeste, las respuestas de Lope al soneto “Patos del aguachirle”; en naranja, los grandes textos del decenio de 1630.

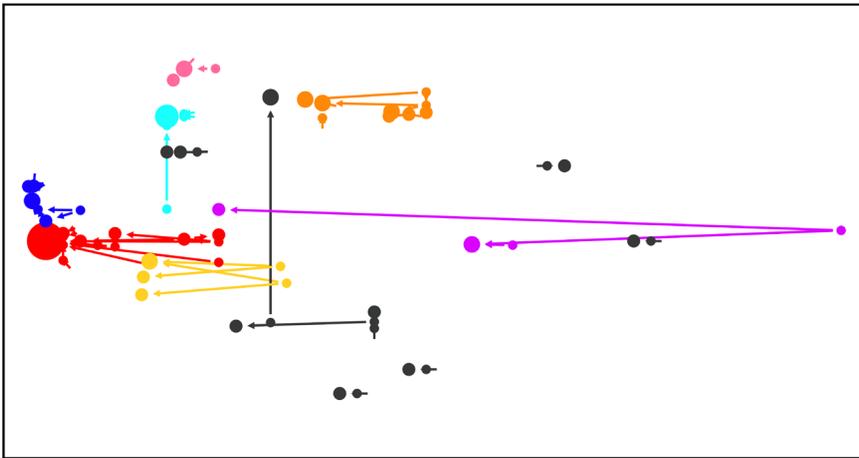


Fig. 11. “Intertextualidad directa en la polémica en función del tiempo”,
apud Elvira y Plagnard, 2021, p. 58.

58. El ya mencionado *Antiantídoto*, in Blanco, Elvira y Plagnard (2021: núm. 27, 572).

59. La *Apología*, in Blanco, Elvira y Plagnard (2021: núm. 28, 572).

60. Elvira y Plagnard (2021: 53).

Conclusión

Sometiéndonos a un proceso de auto-análisis a posteriori, hemos tratado de extraer algunas enseñanzas del proyecto Πólemos —en el que llevamos trabajando unos diez años—, con la esperanza de que puedan aplicarse a otros casos. Vimos que su programa ha consistido y consiste en la edición y estudio de la polémica gongorina, y en general, de los documentos de recepción de Góngora en su siglo. Esperamos que haya quedado claro por qué este material se prestaba a un tratamiento digital y por qué hubiera sido difícil no solo realizar el proyecto por los medios tradicionales, sino incluso concebirlo.

En base a ello, propusimos resaltar la naturaleza de la edición digital, que no es una transformación optimizada de la edición filológica, sino una creación de objetos nuevos a partir de los textos. Con la experiencia limitada de Πólemos y con un par de ejemplos más, ya se puede intuir que estos objetos renuevan nuestra atención por la historicidad de los textos, tanto en su génesis (tradición material), como en su recepción e interpretación (caso de nuestra polémica). A su manera, la construcción de tales objetos textuales implica una renovación y redefinición de la filología.

En el caso de nuestro proyecto, la herramienta clave para la edición digital fue el lenguaje XML-TEI. Su aprendizaje y uso fue la condición para realizar la edición misma, en cuanto operación que se compromete a actualizar de modo óptimo un corpus, poniendo a disposición de quienes lo consulten los medios de su comprensión y contextualización. Por otro lado, nos permitió, mediante métodos como los escenarios de transformación, construir herramientas para leer el corpus en su totalidad, sin hipótesis previas. En especial, construimos una serie de redes de relaciones entre textos y entre agentes de la polémica que constituyen, como lo sugiere el OBVIL (Observatorio de la Vida Literaria), la vida observable de la literatura.

Bibliografía

- ALLÉS-TORRENT, Susanna, “Edición digital y algunas tecnologías aliadas: codificación TEI y transformaciones XSLT”, *Ínsula. Revista de Letras y Ciencias Humanas*, 822 (junio de 2015), pp. 1-30.
- ALLÉS-TORRENT, Susanna, “Crítica textual y edición digital o ¿dónde está la crítica en las ediciones digitales”, *Studia aurea*, 14, 2020, pp. 63-98.
- ANGULO Y PULGAR, Martín de, *Epístolas satisfactorias*, ed. Juan Manuel Daza, Paris, Sorbonne Université – Labex OBVIL, 2018 https://obvil.sorbonne-universite.fr/corpus/gongora/1635_epistolos, https://github.com/gongora-digital/polemos/blob/gh-pages/1635_epistolos.xml.
- ARTIGAS, Miguel, “Un opúsculo inédito de Lope de Vega. El Anti-Jáuregui del Liz. D. Luis de la Carrera”, *Boletín de la Real Academia Española*, XII, 1925, pp. 587-605.
- BLANCO, Mercedes, “La polémica como fermento creativo en el Lope de la vejez”, *Anuario Lope de Vega*, n°14 (2008), pp. 37-66.
- BLANCO, Mercedes, ELVIRA, Muriel, y PLAGNARD, Aude “La recepción polémica de la poesía gongorina: propuesta de catalogación (1612-1692)”, *El universo de una polémica. Góngora y la cultura del siglo XVII*, coords. Mercedes Blanco y Aude Plagnard, Madrid, Iberoamericana-Vervuert, 2021, pp. 557-680.
- BLANCO, Mercedes y MULAS, Margherita, *Carta sobre la poesía nueva de don Luis de Góngora*, Paris, Sorbonne Université – Labex OBVIL, 2018 https://obvil.sorbonne-universite.fr/corpus/gongora/1634_cartas-cascales, https://github.com/gongoradigital/polemos/blob/gh-pages/1634_cartas-cascales.xml.
- CACHO CASAL, Rodrigo, “¿Qué captas, nocturnal, en tus canciones...?. Edición y estudio de un soneto antigongorino de Quevedo”, *Calíope*, 10.2, 2004, pp. 51-71.
- CARMONA, Alba, “¿Escribía Lope de Vega con falsilla? Aproximación a *El mayor-domo de la duquesa de Amalfi* y *El perro del hortelano* desde el análisis de las redes sociales”, *Bulletin of Hispanic Studies*, XCVIII/6 (2021), pp. 839-856.
- CARREIRA, Antonio, “Presencia de Góngora en la poesía de Quevedo”, en *El universo de Góngora. Orígenes, textos y representaciones*, ed. Joaquín Roses, Córdoba, Diputación de Córdoba, 2014, pp. 473-494.
- CONDE PARRADO, Pedro, “Antagonistas de Góngora: Quevedo”, en *El universo de una polémica. Góngora y la cultura española del siglo XVII*, eds. Mercedes Blanco y Aude Plagnard, Madrid, Iberoamericana, 2021, pp. 201-214.
- DÍAZ DE RIVAS, Pedro, *Anotaciones a la Segunda Soledad*, eds. Melchora Romanos y Patricia Festini, Paris, Sorbonne Université – Labex OBVIL, 2017, https://obvil.sorbonne-universite.fr/corpus/gongora/1617_soledad-segunda-diaz.
- ELVIRA, Muriel y PLAGNARD, Aude, “Analizar y cartografiar la polémica gongorina: base de datos, catálogo y análisis de redes”, *El universo de una polémica. Góngora y la cultura española del siglo XVII*, coords. Mercedes Blanco y Aude Plagnard, Madrid, Iberoamericana, 2021, pp. 29-64.

- ESPINOSA, Juan de, *Carta... a Juan de Arguijo*, ed. Héctor Ruiz Soto, Paris, Sorbonne Université – Labex OBVIL, 2021, https://obvil.sorbonne-universite.fr/corpus/gongora/1613_carta-espinoza,
- FARIA E SOUSA, Manuel de, *Noches claras* (fragmentos), ed. Aude Plagnard, Paris, Sorbonne Université – Labex OBVIL, 2019, http://obvil.sorbonne-universite.site/corpus/gongora/1624_nochesclaras, https://github.com/gongoradigital/polemos/blob/gh-pages/1624_nochesclaras.xml.
- FERNÁNDEZ DE CÓRDOBA, Francisco, *Examen del Antídoto o Apología por las Soledades de don Luis de Góngora contra el autor del Antídoto*, ed. Matteo Mancinelli, Paris, Sorbonne Université – Labex OBVIL, 2019, https://obvil.sorbonne-universite.fr/corpus/gongora/1617_examen, http://gongoradigital.github.io/polemos/1617_examen.xml.
- GARCÍA AGUILAR, Ignacio, “Una nueva herramienta para el estudio de los paratextos de Lope mediante la edición digital y el etiquetado semántico TEI: PRESOLO”, *Anuario Lope de Vega. Texto, literatura, cultura*, XXIX (2023), pp. 224-286.
- GLORIEUX, Frédéric, “Odette”, Sorbonne Université – Labex OBVIL, 2015, <https://obvil.sorbonne-universite.fr/developpements/odette>.
- GÓNGORA, Luis de, *Poesía*, ed. Antonio Carreira, Sorbonne Université — Labex OBVIL, 2016, https://obvil.sorbonne-universite.fr/corpus/gongora/gongora_obra-poetica, https://github.com/gongoradigital/gongoraobra/blob/gh-pages/gongora_obra-poetica.xml.
- GUTIÉRREZ VALENCIA, Cristina, “Estos días lo he pasado mal con los de la nueva poesía, Lope como antagonista de Góngora” en, *El universo de una polémica. Góngora y la cultura española del siglo XVII*, coords. Mercedes Blanco y Aude Plagnard Madrid, Iberoamericana, 2021, pp. 215-238.
- JÁUREGUI, Juan de, “Carta del licenciado Claros de la Plaza al maestro Lisarte de la Llana” en *Sales españolas o agudezas del ingenio nacional (Segunda Serie)*, ed. Paz y Meliá, Madrid, Sucesores de Rivadeneyra, 1902, pp. 277-296.
- JÁUREGUI, Juan de, *Discurso poético* [1624], ed. Mercedes Blanco, Paris, Sorbonne Université – Labex OBVIL, 2016, http://obvil.sorbonne-universite.site/corpus/gongora/1624_discurso-poetico, https://github.com/gongoradigital/polemos/blob/gh-pages/1624_discurso-poetico.xml.
- LESCASSE, Marie-Églantine, “L’emploi du concept de *propiedad* dans la polémique gongorine”, *Nouvelles méthodes pour une nouvelle poésie : Góngora et les humanités numériques*, Mercedes Blanco (éd.), e-Spania, n° 29, février 2018, <https://doi.org/10.4000/e-spania.27486>
- MATAS CABALLERO, Juan, *Juan de Jáuregui: poesía y poética*, Sevilla, Diputación Provincial de Sevilla, 1990.
- MOLANES, Mónica y Candelas, Manuel Ángel. “Las adaptaciones lopescas de Juan Mayorga: *La dama boba* en el siglo XXI”. *Anuario Lope de Vega. Texto, literatura, cultura*, 2011, Vol. 17, pp. 66-84, <https://raco.cat/index.php/anuariolopedevega/article/view/286225>.

- MONTERO, Juan, “La rivalidad literaria entre Lope y Jáuregui”, *Anuario Lope de Vega*, vol. 14, núm. 2, 2008, pp. 181-212.
- PEZZINI, Sara y RUIZ SOTO, Héctor, “Editar a Gongora con instrumentos digitales: bases metodológicas y teóricas para un hipertexto de la polémica gongorina”, *Revista de Humanidades Digitales*, nº1, 2017, pp. 35-56, <https://revistas.uned.es/index.php/RHD/issue/view/1096>.
- PIERAZZO, Elena, “Modelling Digital Scholarly Editing: From Plato to Heraclitus”, en *Digital Scholarly Editing: Theories and Practices*, ed. Elena Pierazzo y Matthew James Discroll, Open Book Publishers, 2016, pp. 41-58. <<https://doi.org/10.11647/OBP.0095.03>>.
- PLAGNARD, Aude y RUIZ SOTO, Héctor, “Polemos2nodes/polemos2edges : première lecture de la polémique gongorine par l’analyse de réseau”, *Nouvelles méthodes pour une nouvelle poésie : Góngora et les humanités numériques*, Mercedes Blanco (éd.), e-Spania, nº 29, février 2018, <https://doi.org/10.4000/e-spania.27470>.
- PRESOTTO, Marco, “Hacia un modelo de edición crítica digital del teatro de Lope”, *Anuario Lope de Vega. Texto, literatura, cultura*, XXI (2015), pp.79-94.
- ROJAS CASTRO, Antonio, Luís de Góngora, *Soledades*. Edición crítica digital, ed. Antonio Rojas Castro, <<http://soledades.uni-koeln.de/>>.
- ROSES LOZANO, Joaquín, *Una poética de la oscuridad. La recepción crítica de las “Soledades” en el siglo XVII*, Londres, Tamesis Books, 1994.
- SAHLE, Patrick, “What Is a Scholarly Digital Edition”, en *Digital Scholarly Editing: Theories and Practices*, ed. Matthew James Discroll y Elena Pierazzo, Open Book Publishers, 2016, <<https://doi.org/10.11647/OBP.0095>>.
- SLIWA, Krysztof, *Cartas, documentos y escrituras de Luis de Góngora y Argote (1561-1627) y de sus parientes*, Córdoba, Universidad de Córdoba – Ayuntamiento de Córdoba, 2004, I, pp. 322-336.
- , *Cartas, Documentos y Escrituras del Dr. Frey Lope Félix de Vega Carpio (1562-1635)*, Newark, Delaware, Juan de la Cuesta, 2007, I, pp. 112-124.
- VEGA, Lope de, *La dama boba: edición crítica y archivo digital*. Bajo la dirección de Marco Presotto y con la colaboración de Sònia Boadas, Eugenio Maggi y Aurèlia Pessarrodona. PROLOPE, Barcelona; Alma Mater Studiorum – Università di Bologna, CRR-MM, Bologna, 2015.
- VEGA, Lope de (atribuido), *Anti-Jáuregui*, eds. Juan Montero y Cipriano López Lorenzo, Paris, Sorbonne Université – Labex OBVIL, 2023, https://obvil.sorbonne-universite.fr/corpus/gongora/1624_anti-jauregui, https://github.com/gongoradigital/polemos/blob/gh-pages/1624_anti-jauregui.xml.
- ZARCO CUEVAS, Julian, “Las contiendas literarias en el siglo XVII. – III. Una réplica de Lope de Vega contra don Juan de Jáuregui”, *La ciudad de Dios*, CXLIII, núm. 1260, 1925, pp. 272-290.

La edición crítica digital de la obra poética de Garcilaso de la Vega

Eugenia Fosalba

Universitat de Girona
eugenia.fosalba@udg.edu
ORCID: 0000-0002-8204-8563

Susanna Allés-Torrent

University of Miami
susanna_alles@miami.edu
ORCID: 0000-0002-3616-2285

Resumen

Pronapoli es un proyecto con dos objetivos primordiales: por un lado, estudiar la estancia italiana del poeta Garcilaso de la Vega, llevada a cabo en Nápoles a intervalos, desde el verano de 1529 hasta la primavera de 1536; por el otro, publicar una nueva edición crítica digital de su obra poética. Este capítulo aborda las fases generales por las que ha transcurrido la investigación del equipo; a continuación, se presenta la plataforma web que se concibe simultáneamente como instrumento de difusión de los resultados científicos y una herramienta de trabajo; y en fin, y como tema central, se ofrece una visión genérica de la edición crítica digital de la obra poética garcilasiana en curso de elaboración y publicación, reflexionando sobre el proceso ecdótico, el flujo digital filológico, y la presentación final en línea.

Palabras clave

Garcilaso de la Vega; edición crítica digital; modelado de datos, Pronapoli.

Abstract

The Digital Critical Edition of the Poetic Works of Garcilaso de la Vega.

Pronapoli is a project with two main objectives: firstly, to delve into the Italian stay of the poet Garcilaso de la Vega, spanning from the summer of 1529 to the spring of 1536 in Naples; secondly, to release a novel digital critical edition of his poetic oeuvre. This chapter elucidates the overarching phases traversed by the research team; subsequently, it introduces the web platform, conceived concurrently as a conduit for disseminating scientific findings and as a workspace; lastly, it offers a comprehensive outlook on the ongoing digital critical edition of Garcilaso's poetry, delving into the ecdotic process, philological digital flow, and ultimate online presentation.

Keywords

Garcilaso de la Vega; scholarly digital edition; data modeling, Pronapoli.

1. Garcilaso de la Vega y la estancia italiana: nuevos puntos de partida y el proyecto Pronapoli

Aunque Garcilaso sea sin duda uno de los autores más representativos del canon del Siglo de Oro, y que por consiguiente cuenta con una muy extensa bibliografía, hay todavía amplias lagunas de conocimiento acerca de su vida y obra. Estas carencias crecen en torno a su estancia italiana, definitiva en el cosmopolitismo que va adquiriendo su poesía, y que se prolongó, a intervalos, desde el verano de 1529, y de forma más permanente, a partir de agosto de 1532, hasta la primavera de 1536, poco antes de su muerte en La Provenza.

Con el fin de estudiar la estancia italiana de Garcilaso surgió, así, hacia 2016, la iniciativa Pronapoli,¹ en cuyo seno se aglutina un equipo de expertos, procedente de tradiciones filológicas y literarias distintas, que aspira a llevar a cabo, desde una óptica interdisciplinar, un acercamiento a la obra de Garcilaso de la Vega a partir de la recuperación de las coordenadas históricas y literarias de la estancia del gran poeta en Italia. Desde Pronapoli se ha profundizado durante los últimos años, en la reconstrucción de los círculos literarios que frecuentó el poeta, desenterrando academias literarias que permanecían en el olvido, resituando en el tiempo la composición de obras publicadas decenios más tarde (como el *De Poeta* de Minturno), revisando las relaciones de Garcilaso con otros humanistas a través de deudas literarias (mutuas, bidireccionales), nunca atendidas, con poetas contemporáneos, a través de relaciones epistolares clave, poco o nada analizadas, indagando siempre a través de crónicas de la época y la documentación de archivo. Solo así se va perfilando de forma cabal la figura y la obra de uno de los personajes más significativos y relevantes del panorama literario, cultural y político del siglo XVI.

En este capítulo nos centraremos en las fases generales por las que ha transcurrido la investigación del equipo, prestando especial atención a la plataforma web que alberga los resultados científicos más relevantes, y sirve a su vez de nodo de información para la comunidad científica, facilita el camino a los futuros investigadores, y funciona como espacio colaborativo de trabajo. Además, ofrecemos aquí una visión de conjunto de la edición crítica digital de la obra poética garcilasiana que se encuentra en curso de elaboración y publicación, y reflexionamos sobre el proceso ecdótico, el flujo digital filológico, y el resultado final en línea.

1.1. *Garcilaso de la Vega en Italia: Estancia en Nápoles (Fase 1)*

Algunos esfuerzos que venían ya llevándose a cabo desde la última década de los 2000, confluyeron en el 2016 en la primera financiación del proyecto “Garcilaso

1. Pronapoli, <https://pronapoli.com/>

de la Vega en Italia: Estancia en Nápoles”.² Un equipo internacional procedente de filologías latina, italiana e hispánica conformaron la base de un proyecto que desde el principio se presentaba con vocación interdisciplinar (colaborando las distintas filologías española, italiana, clásica, así como uno de los mejores historiadores especialistas en el período napolitano e italiano en general).³ En esta primera ronda de trabajo colectivo, se analizaron de primera mano los textos literarios y epistolares escritos o dirigidos a Garcilaso, así como la bibliografía secundaria publicada recientemente tanto sobre el propio Garcilaso como sobre los poetas de su entorno cultural napolitano. El objetivo consistió en revisar algunos matices de sus presupuestos y abrir nuevas vías de búsqueda en los intersticios de las disciplinas, indagando en las relaciones personales para determinar las mutuas influencias, los canales en que llegaron las posibles fuentes a manos del poeta, y ajustar mejor así el proceso de creación y su timing.⁴

A lo largo de estos años, ha habido tres líneas de trabajo que se alimentan y complementan al paso de nuevas investigaciones, con su pertinente reflejo directo en la plataforma en línea colectiva:⁵

- a. Con el fin de reconstruir el ambiente cultural que rodeó a Garcilaso, se creó una Galería de autores con fichas identificativas de poetas y humanistas. Estos personajes, en algunas ocasiones, son autores de referencia que residieron en Nápoles en tiempos de nuestro poeta, o que precedieron su estancia, como es el caso de Pontano o Sannazaro; en otros, se trata de autores a quienes nuestro poeta trató e intercambió con ellos impresiones y experiencias poéticas, como con Antonio Sebastiano Minturno, Bernardo Tasso o Luigi Tansillo. Así, se han publicado varias fichas bio-bibliográficas gracias a los datos que han ido surgiendo al paso de las nuevas precisiones en nuestros trabajos: Pietro Bembo, Juan Boscán, Vittoria Colonna, Onorato Fascitelli, Miquel Mai, Antonio Sebas-

2. Proyecto FFI2015-65093-P (2016-2020), financiado por el Ministerio de Economía y Competitividad. Los resultados de esta primera fase pueden verse en: <https://pronapoli.com/>

3. Véase la sección Equipo en la web, <https://pronapoli.com/equipo/>

4. Véanse algunos de estos resultados bajo forma de monografías colectivas e individuales: *La égloga renacentista en el reino de Nápoles*, ed. E. Fosalba y G. de la Torre, Bulletin Hispanique, Presses Universitaires de Bordeaux, 2017; *Contexto latino y vulgar de Garcilaso en Nápoles: redes de relaciones de humanistas y poetas (manuscritos, cartas, academias)*, Peter Lang Verlag, 2018; “*Di qui Spagna et Italia han mostro chiaro l'onor*”. *Estudios dedicados a Tobia R. Toscano sobre Nápoles en tiempos de Garcilaso*, Studia Aurea Monográfica, Publicaciones Universitat Autònoma de Barcelona, 2019; *Clasicismo horaciano en tiempos de Garcilaso de la Vega*, ed. E. Fosalba, Bulletin Hispanique, 125-1, 2023.

5. La plataforma de Pronapoli, en su estado actual, se ha construido con uno de los CMS más utilizados hasta el momento, Wordpress, <https://wordpress.org/>. A partir de una plantilla genérica se ha ido adaptado el diseño para responder así a las necesidades de presentación concebidas por el equipo.

tiano Minturno, Luigi Tansillo, Bernardo Tasso, Ludovico Ariosto, Jacopo Sannazaro, Baldassare Castiglione, Alfonso D'Avalos (en proceso), son solo algunos nombres de entre los más conocidos. En la mayoría de las fichas, además de recoger la información más destacada de fuentes ya canónicas como el *Dizionario Biografico degli Italiani*, se allegan datos dispersos en bibliografía más o menos remota, y se van corrigiendo, ampliando y actualizando otros datos.⁶

- b. A la reconstrucción cultural, se ha sumado una sección sobre las Academias, espacios de reunión de humanistas que seguramente debió frecuentar Garcilaso. Con este objetivo, se han señalado una serie de localizaciones que ayudan a dar una idea de la variedad de cenáculos y tertulias, más o menos formales, que en esos años reunían a los poetas de la época.⁷ De esta manera, se ha ofrecido información que había quedado sólo consignada en nuestras publicaciones de forma, a veces, algo marginal, como por ejemplo en notas a pie de página, mientras que el geolocalizador visualiza la ubicación precisa de las mismas.
- c. Una tercera sección corresponde a la Biblioteca digital que tiene por objetivo la reconstrucción de las reproducciones o referencias a los textos manuscritos que hasta la fecha eran desconocidos o simplemente resultaban poco accesibles, mientras que ahora se ofrecen en un único espacio virtual.⁸ Por ejemplo, se han editado y traducido los epigramas griegos de Giovanni Andrea Gesualdo, pariente de Minturno y autor de una edición comentada del *Canzoniere* de Petrarca que se publicó en Venecia en 1533. Así, en la mayoría de los casos ofrecemos reproducciones, traducciones y transcripciones de manuscritos, de textos hasta ahora inéditos y estudiados por primera vez, pertenecientes al volumen facticio (Biblioteca Nacional de Nápoles XIII AA 63), que hasta hace poco había sido el principal testimonio de dos de las odas latinas de Garcilaso de la Vega.⁹

Estas tres líneas de trabajo —con sección propia en la misma web del proyecto (*Galería de autores, Academias, Biblioteca digital*)— han corrido de forma paralela a otros esfuerzos destinados a recoger diferentes fuentes primarias correspondientes a ediciones de obras de autores italianos de los primeros cuarenta años del siglo XVI. Además, se ha recopilado bibliografía secundaria, de aportaciones de la crítica tanto hispanista como italianista, útil para llevar a cabo la investigación en torno a Garcilaso y clasificada por materias y formatos (artícu-

6. La Galería de autores puede consultarse en, <https://pronapoli.com/autores/>

7. Para la creación del mapa, se utilizó Google Maps, y se ha añadido a la plataforma de Wordpress, <https://pronapoli.com/academias/>.

8. La Biblioteca digital se encuentra disponible en: <https://pronapoli.com/biblioteca-digital/>

9. Para una descripción detallada de este manuscrito, véase el apéndice en Fosalba (2018: 297-328).

los, capítulos, libros, etc.). En fin, se ha añadido también un Catálogo de impresos en el *Regno di Napoli* aparecidos también durante la primera mitad del siglo XVI.¹⁰

A lo largo de estos primeros años, además, el equipo ha ido publicando diferentes volúmenes relativos a la estancia italiana de Garcilaso, así como un listado de *Centros de documentación* donde los miembros del equipo desarrollan sus investigaciones. Estos trabajos preliminares proporcionaron un bagaje sobre el contexto napolitano en el seno del cual Garcilaso desarrolló la parte más relevante de su obra. Asimismo, esta labor previa nos permitió empezar a ahondar en aspectos clave que constituirán las fases sucesivas del proyecto, como el estudio del horacianismo por el que Garcilaso fue tan valorado en Nápoles, y el plan de edición de las obras de corte más o menos horaciano (latinas y castellanas) extrayendo el máximo partido de las posibilidades que ofrece la tecnología digital.

1.2. *Garcilaso de la Vega en Italia: Clasicismo horaciano (Fase 2)*

En la segunda fase de financiación del proyecto nos propusimos un doble objetivo:¹¹ por un lado, profundizar en el estudio de su neoclasicismo horaciano, latino y vulgar, en el contexto italiano y napolitano de Garcilaso, y, por el otro, en estrecha vinculación con lo anterior, editar digitalmente las odas latinas y las obras de este sesgo en castellano. Fueron afirmaciones sobre la poesía de Garcilaso como las de Pietro Bembo, seguidas por Paolo Giovio y Girolamo Seripando acerca de la excelencia del horacianismo de sus odas, las que decidieron el rumbo de esta segunda edición del proyecto, a medida que fue surgiendo la pregunta acerca del porqué de este elogio compartido, a la que se trata de contestar desde el estudio y edición de la obra latina, y también castellana, escrita bajo el influjo del venusino.

Pero no solo nos centramos en el estudio filológico: el sistema de investigación que empleamos en Pronapoli trata de poner en relación todos los aspectos del contexto (literarios, históricos, acerca de las relaciones interpersonales descubiertas en poemas y en epistolarios o en los archivos), para reconstruir la red intelectual en la que el poeta desarrollaba su obra. Esta nueva senda investigadora, que avanza de forma radial, ya ha dado numerosos frutos y se promete fecunda y extensible a otros tantos poetas siglodoristas que

10. Toda esta información bibliográfica está de momento en la base de datos del Wordpress, pero en un futuro se espera que pueda exportarse a un gestor bibliográfico, como Zotero, para poder compartir más fácilmente los metadatos y facilitar la creación de bibliografías específicas.

11. Proyecto PID2019-107928GB-I00 (2020-2024), financiado el Ministerio de Ciencia, Innovación y Universidades (MCIU).

residieron largas temporadas en Italia o Europa y que merecerían otros tantos proyectos al estilo de Pronapoli. Baste recordar a Juan del Encina, Torres Naharro, Cristóbal de Castillejo, Gutierre Cetina, Hernando de Acuña, Jerónimo de Urrea, Diego Hurtado de Mendoza, Jorge de Montemayor, Francisco de Aldana, Luis Gálvez de Montalvo, Miguel de Cervantes, Juan de Tassis y Peralta, Francisco de Quevedo, entre tantos otros. Se trata de superar los vetustos límites establecidos por los compartimentos estancos de las actuales disciplinas, que separa las filologías en autores que componen en latín o en vulgar, y que aparta la historia y la biografía de las circunstancias de la composición. En definitiva, tratamos de descubrir lo que ha quedado en ángulos muertos y que arroja una luz nueva, muy reveladora, sobre la interpretación de la obra en relación con las vicisitudes de su autor.

La consecuencia lógica de la primera etapa de recolección y análisis de datos, así como publicación de los resultados de las primeras vinculaciones entre ellos, fue la necesidad de una nueva edición crítica y anotada que, por el volumen de trabajo y la gran cantidad de materiales paratextuales, así como por el equipo multidisciplinar que implicaba, debía ser concebida, creada y publicada en formato digital y de manera colaborativa. Se decidió así recopilar los textos que participan con mayor o menor intensidad de esta influencia capital de Horacio, compuestos en la última etapa de la vida de nuestro poeta, para ofrecer una renovada edición y estudio de los mismos. Entre los textos seleccionados figura la compleja *Égloga II* (con rasgos típicamente horacianos solo en la estancia dedicada al *beatus ille*), y otras obras de factura más decididamente horaciana, como la *Epístola a Boscán*, la *Elegía II*, y la *Ode ad Florem Gnidi*. Por lo que respecta a sus odas latinas, se editan las tres odas conocidas del poeta, I *Ad Antonium Thylesium Ode*; II *Garsiae Lasi Ode ad Genesium Sepulvedam*, y III *Sedes ad Cyprias Venus*. Y a estas se han sumado dos más que han aparecido por el camino, redescubiertas por María Czepiel, miembro del equipo Pronapoli; obras relevantísimas, antes desconocidas, que, además, están transcritas en un nuevo testimonio, de factura alcalaína, muy valioso para las demás odas latinas ya conocidas, que también transcribe. Las nuevas odas son las siguientes: IV *Intentos humilis, Bembe, sonus lyrae*, y V *Brasicane, meis iure sodalibus*.¹² Este carácter multidisciplinar de nuestras indagaciones ha beneficiado también en gran medida la anotación de las fuentes en su variada extensión de influencias clásicas, neolatinas e italianas, de obras que están en el centro de ese laboratorio de experimentación que fue la Nápoles renacentista y, a la vez, se han podido fijar los textos latinos, en una edición por primera vez crítica, y muy ampliamente anotada, de los mismos.

12. Manuscrito conservado en la Biblioteca Nacional de la República Checa, con la signatura E IX 1 (Czepiel 2022).

1.3. *Estudio y edición de la obra poética castellana de Garcilaso de la Vega (Fase 3)*

La siguiente fase del proyecto todavía en curso de elaboración (previsiblemente 2024-2028) abarca el estudio y la edición de la obra poética castellana producida por Garcilaso, proponiendo tres líneas de acción. En primer lugar, se pretende ampliar el radio de las investigaciones, incluyendo un estudio más pormenorizado de las fuentes, especialmente gracias al enriquecimiento de la anotación de la obra castellana que incluirá los estudios llevados a cabo sobre las fuentes clásicas e italianas. En segundo lugar, se continuará con la edición crítica digital cubriendo ahora el resto de la obra castellana. Tal empresa no está exenta de desafíos ecdóticos, pues esta presenta notables problemas textuales aún no revisados (como se verá en el siguiente apartado). El resultado ampliará la edición crítica que hasta ahora se ha circunscrito a las cinco odas latinas y parte de la obra castellana seleccionada por su horacianismo, por lo que comprenderá el resto de la obra castellana: las ocho coplas castellanas, los cuarenta sonetos, las cuatro canciones petrarquistas, las églogas I y III, y la Elegía I. En fin, también está prevista la redacción de un nuevo estudio de la biografía del poeta, por parte de Fosalba, con la colaboración de Adalid Nievas, teniendo en cuenta trabajos publicados que han arrojado nueva luz sobre el sentido de varios de los poemas clave, vulgares y latinos, del gran poeta (*Oda a Ginés de Sepúlveda*, *Ode ad Florem Gnidi*, *Égloga II*, entre otros).

2. ¿Por qué una nueva edición digital de la obra poética de Garcilaso de la Vega?

En 2021, Joseph Reed, investigador de la Universidad de Brown, perteneciente al proyecto Pronapoli, ya anunció en *Studia Aurea*, que hacía falta volver sobre estos textos, en una edición por fin crítica, al tiempo que planteó posibles alternativas a las lecturas aceptadas. Ahora, con el nuevo testimonio para las tres odas aportado por Czepiel, se han podido revisar los textos con una nueva mirada. El descubrimiento de Czepiel, relevantísimo, de un nuevo testimonio manuscrito, conservado entre los últimos folios en blanco de un ejemplar de *Doctissimorum nostra aetate Itolorum epigrammata*, París, Nicolas le Riche, 1547, aporta las tres odas conocidas en una nueva copia, y dos nuevas odas latinas, una dedicada a Pietro Bembo y otra a Johan Alexander Brasicanus. El libro, editado por Jean de Gagny, es una antología en verso de composiciones de poetas italianos de finales del siglo xv, como Flaminio, Molza, Navagero, Cotta, Lampridius y Sadoletto, y dicho ejemplar se encuentra en la Biblioteca Nacional de la República Checa. En las últimas hojas del libro, una mano probablemente española del siglo xvi ha escrito otra selección de poemas de Sadoletto, Garcilaso, André de Resende, Jaime Juan Falcó, Die-

go Hurtado de Mendoza, Pietro Gravina, Juan Ramírez, Honorato Juan y Johannes Secundus. Estos españoles cuyas obras se copian suelen ser posteriores a los poetas impresos en la antología; Czepiel también sospecha que el copista, o al menos su fuente, procedía de los círculos universitarios de Alcalá. Por ejemplo, una nota manuscrita al final del libro, aparentemente de la misma mano que la antología poética a la que sigue, da cuenta de cómo Andreas Navagero estuvo en Alcalá de Henares con el doctor Ramírez (es decir, Juan Ramírez, catedrático de retórica de la universidad). Además, el nuevo testimonio (D) aporta algunas lecturas más fiables que muestran que no derivan del manuscrito napolitano (BNN XIII AA 63) como los testimonios ya conocidos de *Ad Antonium Thylesium* y *Sedes ad Cyprias*. En particular, el nuevo testimonio le permite recuperar la línea 53 de *Sedes ad Cyprias*.¹³ No obstante, la inclusión de lecturas que necesitan enmienda y que son comunes a XIII AA 63 *Mx* sugiere que ambas tienen una fuente común, como sugiere Czepiel: “In particular, the new witness allows us to recover line 53 of ‘Sedes ad Cyprias’, missing until now (unfortunately disproving my own earlier conjecture of what the line could have contained). Nevertheless, the inclusion of readings which need emendation and are common to *Mx* suggests that both have a common source” (Czepiel 2022: 746). Joseph Reed, en su edición de la Oda I, por ejemplo, también observa variantes de autor que refuerzan el lado platónico de la oda.

Urgía asimismo compartir en formato abierto los resultados de esta primera edición crítica de las odas latinas en la historia de la filología española, con las dos nuevas incorporadas al elenco neolatino, en un formato digital, que reflejara sobre la pantalla los avances esenciales de la crítica textual digital, y que permitiera dejar a la vista las aportaciones colectivas sobre cada obra, aunque la responsabilidad sobre cada una de ellas recayera sobre un investigador en especial (R. Lokaj (Ode I), C. Corfiati (Ode II), J. Reed (Ode III), M. Czepiel (Odes IV and V)). Joseph Reed, vale también la pena mencionar aquí, se encargó de acabar de pulir la fijación de la totalidad de los cinco textos latinos.

En lo que concierne a la obra poética garcilasiana escrita en castellano, esta presenta desafíos de mayor alcance todavía. Se ha detectado la necesidad de llevar a cabo un cotejo exhaustivo de todos los ejemplares conservados de las primeras ediciones, puesto que una labor así no se ha llevado a cabo desde Keniston (1925), y aun así tampoco en este caso resulta completa, además de que presenta un aparato negativo. Se hace a su vez imprescindible ofrecer un

13. Véase la transcripción del manuscrito de la BNN en Pronapoli, con la marca del verso 53 que falta por medio de puntos discontinuos, ausente en el testimonio napolitano, y por supuesto en su copia vaticana, a buen seguro de Angelo Colocci (Vat. Lat. 2836, y no Lat. 2836, como suele transmitirse en las ediciones garcilasianas), <https://pronapoli.com/biblioteca-digital/transcripcion/sedes-ad-cyprias-venus/>.

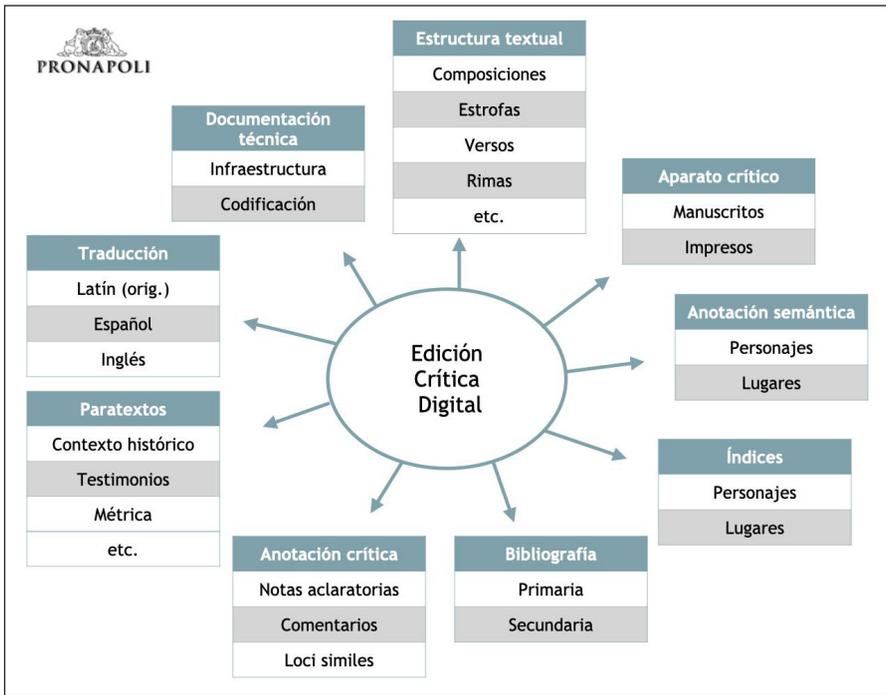
análisis pormenorizado de todos los primeros testimonios para observar los posibles cambios en las distintas emisiones y estados. Al nuevo y exhaustivo cotejo se añadirá un estudio completo de la tradición textual de las ediciones impresas de Garcilaso en las condiciones en que se difundió (tanto en compañía de la obra de Boscán, es decir, tal y como apareció en la *princeps* de 1543, como en edición exenta —a partir de 1569, Salamanca, en la edición llamada de Garcilaso “divorciado de Boscán”—), así como se tendrán presentes los textos que proporcionan las ediciones de los comentaristas. A este estudio se adjuntará una lista de descripciones bibliográficas analíticas también exhaustiva; un análisis pormenorizado que además controlará toda la tradición manuscrita, adjuntando sus consiguientes descripciones, sin dejar de buscar nuevos testimonios, así como, por supuesto, con una colación de primera mano. También resulta imprescindible una anotación que contemple las recientes aportaciones de la crítica, y vuelva sobre la huella de los clásicos grecolatinos y la literatura neolatina del quinientos sin conformarse con las aportaciones de los comentaristas que no bastan ni mucho menos para una anotación completa de las fuentes clásicas de la obra garcilasiana, como se está echando de ver recientemente (Fosalba 2021, 2023).

En cuanto a las razones que nos decidieron a concebir una edición crítica digital, cabe señalar, en primer lugar, que había la necesidad de reconstruir y ofrecer en un único espacio y con criterios ecdóticos uniformes el corpus poético garcilasiano en su integridad, especialmente después de la aparición de las dos composiciones latinas desconocidas hasta la fecha. También nos interesaba idear la edición según el nuevo paradigma digital y siguiendo la idea de archivo donde tuvieran cabida materiales de diferente naturaleza, como reproducciones o paratextos, o diferentes niveles de anotación crítica, que se desplegaran ante un texto limpio y de forma simultánea. Un proyecto así planteaba dudas ante una presentación en papel por obvias cuestiones de espacio y economía. En fin, una edición digital nos abrió las puertas a concebir la posibilidad de ofrecer una edición en varias lenguas europeas.

3. Desiderata filológicos y el modelo para la edición crítica digital

Al inicio de la segunda fase del proyecto se inició el prototipo de la edición digital y para ello se mantuvieron discusiones sobre su naturaleza, acerca de lo que debería contener y cómo era más aconsejable proceder. Una de las cuestiones centrales consistió en establecer qué tipo de edición se deseaba ofrecer: se acordó, y por los motivos aducidos más arriba, que debía tratarse de una edición crítica que revisara de primera mano todos los testimonios disponibles. Otra de las discusiones giró en torno a las diferentes secciones que debíamos incluir, y especialmente sobre el tipo de anotación con la que nos parecía más adecuado acompañar los textos poéticos garcilasianos. Estas y otras cuestiones nos llevaron

a la elaboración de un esquema conceptual del modelo de datos que se explica gráficamente a partir de la Fig. 1.¹⁴



Así pues, la reflexión inicial partió del establecimiento del texto y de la estructura concreta de cada una de las composiciones poéticas (odas, églogas, etc.) con las que debíamos trabajar, así como sus elementos constituyentes (estrofas, versos, rimas). El texto, además, debía ir acompañado de un aparato crítico que reflejara el trabajo filológico del equipo y las nuevas colaciones o decisiones editoriales que se iban llevando a cabo a partir de los testimonios manuscritos y las ediciones antiguas de Garcilaso. En algunas ocasiones, nos hallamos ante un manuscrito único, mientras que en otros, especialmente en el de las ediciones quinientistas, se conservan un número elevado de ediciones.

A continuación, se previó un enriquecimiento semántico del texto que podría dividirse en dos grandes grupos.¹⁵ Por un lado, necesitábamos una anota-

14. Para una definición de modelado o modelización de datos, véase: <https://www.vocabularyserver.com/tadirah/es/index.php?tema=35&/>, y también a este propósito: Allés-Torrent 2019.

15. Al hablar de anotación semántica nos referimos, en este caso, a la codificación en lenguaje XML a partir de las guías directrices de la Text Encoding Initiative que se ha aplicado tanto a los textos poéticos como a los paratextos. Las directrices sobre este sistema de marcado están disponibles en: <https://tei-c.org/>.

ción semántica, en el interior del mismo texto, que aclarara los aspectos fundamentales, tipográficos, nombre de personajes y lugares; por el otro, una anotación que se desplegara en tres niveles conectados directamente con el texto. Así, el texto poético se relaciona con una serie de anotaciones elaboradas por los investigadores que se distinguen entre “Notas”, aclaraciones breves concebidas como apoyo a una primera lectura; “Comentarios”, excursos más eruditos que por lo general reflejan las investigaciones llevadas a cabo por los diferentes miembros del equipo; y los *loci similes*, locuciones o citas directas o indirectas de otros autores, con atención a las referencias clásicas pero también contemporáneas, ya sean en latín o vulgar. Esta última sección resulta imprescindible para estudiar con más profundidad las fuentes de las que se valió Garcilaso en la elaboración de su poesía, teniendo en cuenta que a veces las fuentes son intercambiables, y que no siempre es Garcilaso quien recibe pasivamente el influjo (véase al respecto, Fosalba 2022).

La edición debía ofrecer además dos herramientas básicas que facilitaran la consulta: una opción de búsqueda con texto libre, y otra a partir de índices extraídos a partir de la codificación en XML-TEI que recogieran básicamente los nombres de personajes y los nombres de lugar señalados anteriormente, así también como las referencias bibliográficas.

Otra parte esencial de nuestro modelo fue la elaboración de paratextos que ofrecieran una visión lo más completa posible de cada una de las composiciones poéticas. En primer lugar, se previó una sección dedicada al *Contexto histórico* para entender mejor las circunstancias en las que se gestó el poema y las coordenadas culturales y literarias de la composición. En segundo lugar, se añadió una sección para la *Descripción de los testimonios*, tanto de los manuscritos como de las ediciones cotejadas. Estos detalles codicológicos y bibliográficos se encuentran descritos en el mismo archivo XML-TEI de la sección de descripción de la fuente (sourceDesc, msDesc).¹⁶ El mayor reto, en este sentido, consistió en encontrar un método con el que todo el equipo pudiera trabajar de manera ágil, pues rápidamente nos dimos cuenta que cada investigador tenía una manera diferente de describir las fuentes. A ello se le contraponía la estructura bastante rígida del marcado TEI en la descripción de las fuentes primarias. Por ello, se decidió establecer un formulario, que respondía a la estructura propuesta por las Guías directrices de la TEI, a partir del cual los editores pudieran recoger las informaciones básicas (signatura y datos identificativos de la fuente, la descripción del manuscrito o edición, descripción física, historia de la fuente, origen, procedencia, adquisición, etc.). Una tercera sección debía ofrecer un comentario sobre la métrica, aunque este pudiera tener dimensiones desiguales. Algunos poemas al no presentar innovaciones o particularidades originales métricas se limitan a un breve comentario, mientras que en otros casos, como en el de la

16. Véase el elemento msDesc, <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-msDesc.html>

Ode II, editada por Claudia Corfiati, se despliega un sistema métrico más complejo, el asclepiadeo B, compuesto por estrofas de cuatro versos (dos asclepiadeos menores, un ferecrateo y un gliconeo).

Al tratarse de un equipo internacional, por otro lado, se decidió asimismo que los textos se ofrecerían en diferentes lenguas (por ahora, español, más adelante, en inglés). Desde el punto de vista de la codificación de XML-TEI, algunos elementos complican un marcado multilingüe pues no permiten la repetición del mismo elemento (por ejemplo, si un elemento es repetible, puede distinguirse a través de un atributo `@xml:lang` con la especificación de lengua). Tal es el caso, por ejemplo, del elemento “summary”, en la descripción de las fuentes primarias, que no permite su repetición y nos obliga a recurrir a otro elemento descendiente, como por ejemplo una nota, lo cual no es lo ideal.¹⁷ Ahora bien —y sin querer cuestionar la utilidad de una edición multilingüe—, también podríamos plantearnos por qué los investigadores al afrontar una edición digital asumimos retos tan laboriosos, como es la publicación de todo el contenido en diferentes lenguas, cuando en una edición en papel sería básicamente imposible.

Nuestro modelo requería dos aspectos más. Por un lado, y como no podría ser de otra manera, se debía recopilar y estructurar una serie de bibliografía, tanto primaria como secundaria. Estas referencias se han ido anotando en el mismo texto para poder ser recuperadas en las bibliografías específicas.¹⁸ Por otro lado, debíamos redactar una documentación técnica en la que se describiera la codificación en XML-TEI llevada a cabo, para que futuros integrantes del proyecto puedan efectuar el marcado de manera autónoma.¹⁹

4. Sobre el proceso y el modelo de edición digital

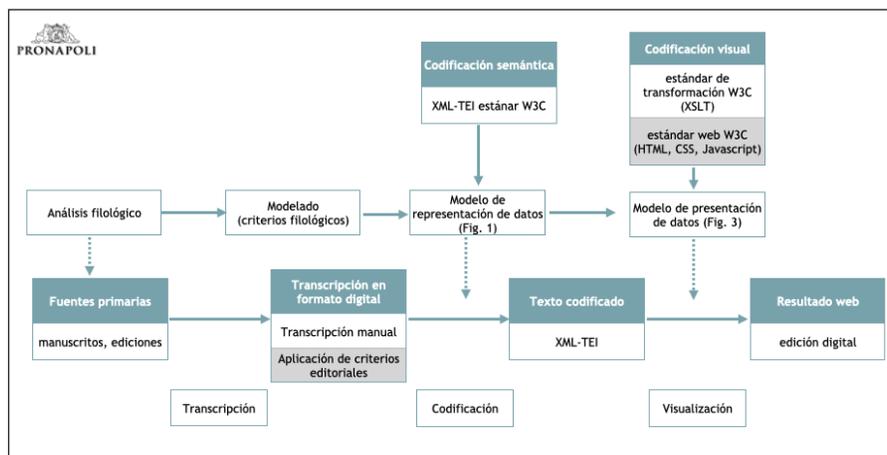
El proceso editorial adoptado para llevar a cabo la obra poética garcilasiana se ha dividido en tres grandes fases: transcripción de la fuente, codificación, y visualización. Para mostrarlo de forma gráfica, ofrecemos un esquema, en la siguiente Fig. 2, del proceso editorial que hemos llevado a cabo, basándonos en el modelo de representación de datos ya propuesto por Rehbein y Fritze (2015).²⁰

17. Para el elemento “summary”, <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-summary.html>

18. El sistema ha consistido en recopilar la bibliografía con el programa Zotero y asignar el identificador que este crea. Así, cada ítem puede ser recuperado con todos sus metadatos a partir de la API de Zotero.

19. Los documentos de cada una de las composiciones poéticas también estarán disponibles en el repositorio de GitHub.

20. Para más detalles y bibliografía sobre la descripción del proceso ecdótico digital, véase Allés-Torrent (2020).



En un primer momento, cada uno de los editores inicia el proceso de análisis filológico y bibliográfico del testimonio base, ya sea que se conserve la composición en uno o varios manuscritos, o bien que forme parte de la tradición textual impresa. Paralelamente, cada editor redacta, con la ayuda del equipo, la serie de paratextos que acompañan la composición y que se añade a las diferentes secciones (contexto, descripción de los testimonios, métrica, y anotaciones en notas, comentarios, y *loci similes*). Para llevar a cabo este análisis, el equipo se basa en el modelo específico de representación de datos presentado arriba (véase la Fig. 1). Una vez el editor filológico ha transcrito la fuente, aplicado los criterios editoriales, y redactado los paratextos, entra en juego el editor “digital” que lleva a cabo la codificación a partir del estándar web XML y siguiendo las directrices de la Text Encoding Initiative.²¹ En esta segunda fase, pues, se procede a la codificación del texto. A continuación, se establece un modelo claro de presentación de datos, es decir, la presentación web final que el equipo ha concebido para la edición. En nuestro caso, un diseñador web construye la interfaz web, mientras que el editor digital se encarga de programar la transformación de los documentos XML-TEI a través de un script en XSLT, otro lenguaje estándar creado por el W3C, adaptándolo a la plantilla web.²² En fin, se genera el resultado como

21. Distinguimos aquí entre estos dos tipos de editores, pues cada uno ha llevado a cabo un trabajo diferente (véase, en este monográfico, la discusión de J. Tronch sobre la terminología relativa al editor). Lo ideal, ciertamente, sería que fuera el mismo editor que realizara tanto el trabajo filológico y de investigación, como la codificación, pero a día de hoy, esto suele suceder pocas veces. Por editor digital, pues, entendemos el que toma el relevo una vez el texto ha sido establecido y procede a su codificación y, en nuestro caso, a la coordinación de la infraestructura y presentación digital.

22. Véanse las especificaciones del lenguaje XSLT, <https://www.w3.org/TR/xslt-30/>, y para el caso específico de las ediciones digitales: Allés-Torrent 2015.

páginas HTML estáticas: cada una de las composiciones corresponde a un documento HTML diferente. Huelga decir, a este respecto, que en términos de infraestructura web para esta edición, se ha recurrido a un concepto de tecnología mínima, utilizando solo lenguajes web estándares y generando el resultado final como páginas HTML estáticas. Así pues, por un lado existen los documentos XML-TEI, por el otro, un transformador escrito en XSLT, y en fin, el resultado en HTML.²³ Esta decisión vino motivada por la experiencia previa llevada a cabo en la sección Biblioteca digital,²⁴ en la que se intentaron incluir los archivos XML-TEI en la infraestructura del CMS utilizado (Wordpress). El resultado constató la complejidad de utilizar herramientas prediseñadas para las ediciones filológicas digitales que presentan problemas de integración y visualización.

En cuanto al diseño web y la experiencia de usuario, nos pareció imprescindible asumir la centralidad del texto original y aspirar a que éste fuera siempre visible (véase la Fig. 3, donde el texto aparece a la izquierda de la pantalla). A partir del texto poético, se ramifica el paratexto a través de una serie de símbolos que conducen a las diferentes secciones. En cuanto al aparato crítico, este planteó el reto de su presentación, un aspecto todavía no resuelto en términos de crítica textual digital. Una imitación en versión digital del aparato tradicional basado en las dos dimensiones del papel, continúa siendo algo críptico para la mayoría de la audiencia y carece de toda aportación al nuevo paradigma digital, mientras que la ideación de un aparato interactivo carga a veces sobremedida la interfaz web. La visualización del aparato crítico permanece uno de los grandes retos. Por el momento, nuestra elección ha consistido en señalar con un símbolo en el texto las palabras o sintagmas con variantes y reflejarlo en la sección de la derecha junto con las otras pestañas y opciones editoriales. Por su lado, la anotación crítica (notas, comentarios, *loci similes*) se agrupan en un símbolo en forma de flecha con un menú desplegable y conducen al usuario a la anotación correspondiente en el panel de la derecha manteniendo siempre la referencia al texto original. En fin, existen algunas secciones que ofrecen textos de dimensión desigual, lo cual puede romper en cierta manera la armonía del formato. Tal es el caso de las secciones dedicadas a la métrica, o a la descripción de testimonios que, aunque el equipo haya trabajado en la homogenización de las descripciones (a partir de un formulario), a veces el editor ha considerado proporcionar más o menos detalles.

23. Todos los archivos estarán disponibles en el repositorio GitHub del proyecto.

24. Véase, <https://pronapoli.com/biblioteca-digital/>



PRONAPOLI

Soledad amena



Intentos humilis, Bembe, sonus lyrae

- 01 Intentos humilis, Bembe, sonus lyrae
- 02 sensus forte tuos si auocart arduis
- 03 ingratum a studiis, parce precor libens.
- 04 non sum, qualis erat, modis
- 05 qui nunc dulcionis flectere pectora
- 06 idem nunc aliter dicere flumine
- 07 verborum rapido facta volubilis
- 08 heroum valuit chey
- 09 nec qualis Latio qui numeros prior
- 10 ostendit fidicen, non sine gloria
- 11 Graios queis renuens ardua dicere
- 12 se furtim volucer supra
- 13 insurgit solido grantor, et modum
- 14 hinc penis staturis dexter ad Aufidii
- 15 ripas se recipit floricolum solum
- 16 demulcens vario sono.
- 17 Non, inquam, numero nobilium deus
- 18 cui cordi cythara est, carmina cui placent,
- 19 vatium me inseruit nec dedit ultimum
- 20 alis aëra scindere.
- 21 O si forte detur tunc ego laudibus
- 22 pennatis agilis tollere me tuis
- 23 ex imo haud dubitem tangere et ardua
- 24 claro vertice sydera,

GARCILASO DE LA VEGA, ODA IV

Contexto
Métrica
Traducción
Locis similes
Notas
Comentarios
AptCrit
Bibliografía

Contexto

Oda horaciana en latín en alabanza de **Pietro Bembo**, humanista, poeta y futuro cardenal italiano. Entre sus obras más famosas se encuentran *Gli asolani* (1497-1504), que constituye un diálogo sobre la naturaleza del amor; *Prose della volgar lingua* (1525), en la que afirmaba que el toscano del siglo XIV debía ser el modelo de la lengua literaria italiana; y una historia de Venecia (véase infra, vv. 3235). También fue interlocutor en *Il Cortigiano* de **Baldassare Castiglione**, traducido por el amigo de Garcilaso, **Juan Boscán**, y para el que **Garcilaso** escribió una epístola prologal.

Es muy probable que, con de este poema, Garcilaso esperara ser admitido en los círculos de la élite intelectual europea, además de obtener reconocimiento como poeta, ganándose con él la aprobación de Bembo (sabemos por una carta de Bembo a Garcilaso que no se habían conocido en persona, véase [Fosalba 2019: 214](#)). De hecho, la poesía panegírica y la inmortalidad poética es el tema de gran parte del poema: mientras Garcilaso afirma que no es tan hábil como los famosos poetas líricos clásicos (vv. 4-20), afirma que si pudiera escribir poesía lírica, lo haría en alabanza de Bembo para poder ganar renombre (vv. 21-28). De hecho, Garcilaso alaba a Bembo por su poesía y por su historia (inédita) de Venecia (vv. 31-35). Incluso afirma que Iberia está celosa de Venecia (vv. 35-40), y la compara con **Alejandro Magno** llorando ante la tumba de **Aquiles** (vv. 41-56).

Se trata de un poema muy metapoético. Garcilaso empieza expresando su "ansia de imitación" (por emplear los términos de Harold Bloom) con una referencia específica a los autores clásicos Píndaro y Horacio, y a su contemporáneo Bembo. [Verg. Ed. 6.3.5](#); y [Prop. 3.3.13-24 Carm. 1.1, 2.20, 3.30](#) o confiesa su deuda con otros antecedentes poéticos (Carm. 4.2, 3.30).

CREDITOS

Garcilaso de la Vega, Oda IV editada por **Maria Czepiel**. En: *Soledad amena*. Edición crítica y digital de la obra poética de Garcilaso de la Vega, dir. E. Fosalba. Pronapoli, 2024. <<https://pronapoli.com/ediciondigital/oda4/>>. [Fecha de consulta]

TEI

Descarga el archivo XML-TEI

5. Conclusiones

Pronapoli ha iniciado una senda productiva de investigaciones en torno a la estancia del gran poeta Garcilaso de la Vega en Italia. Con el fin de reconstruir las coordenadas históricas y literarias de este período, hemos creado una plataforma web, que va más allá de la simple difusión de los resultados y se concibe como una herramienta de trabajo. El uso de un gestor de contenidos permite a los colaboradores subir sus aportaciones, al mismo tiempo que sirve como centro de información para llevar a cabo sus investigaciones. A su vez, hemos puesto a disposición de un público amplio una edición en acceso abierto de las cinco odas latinas y parte de la obra vulgar (más o menos horacianas) de Garcilaso, en un texto fijado tras un nuevo cotejo, mucho más exhaustivo, y unas notas —*loci similes*, notas aclaratorias, comentarios más eruditos extensos—, se conectan entre sí, al texto poético, y a las otras partes de la plataforma Pronapoli. A todo ello se suman la galería de autores, el mapa de las academias, los textos transcritos y rescatados del olvido, la recopilación de fuentes primarias y secundarias online que, en su conjunto, proporcionan materiales para el contexto histórico-cultural en una amplia gama de estratos de la realidad italiana y napolitana de la primera mitad del siglo XVI, y que son imprescindibles para situar y entender con todas las sutilezas y precisiones del caso, la compleja, cosmopolita y extraordinaria obra poética del primer gran poeta moderno de la historia de la literatura española.

Desde el punto de vista de las humanidades digitales en general, y de la práctica de la crítica textual digital en particular, hemos aspirado a ofrecer una edición digital que responda al nuevo paradigma tanto en su teoría y método como en su práctica (Sahle 2016; Allés-Torrent 2020). Resulta claro que la edición de la obra poética de Garcilaso se aleja del formato impreso y se convierte en un amplio depósito de materiales que gravitan alrededor del texto. Así también, el flujo de trabajo, concebido a partir del modelo de representación de datos expuesto más arriba, ha guiado el método filológico de los editores de manera coordinada y paralela. Simultáneamente, se ha concebido una infraestructura web que prevé una estructura de publicación sostenible y escalable para poder integrar otras obras garcilasianas, como esperamos que suceda en la siguiente fase del proyecto, para la obra poética castellana. En fin, la estructura web y todos los contenidos que la conforman, incluidos los archivos fuentes con la codificación en XML-TEI se pondrán a disposición de la comunidad en un repositorio abierto para cumplir con las buenas prácticas del acceso abierto y la transparencia en nuestra práctica científica digital.

Bibliografía

- ALLÉS-TORRENT, Susanna, “Edición digital y algunas tecnologías aliadas: codificación TEI y transformaciones XSLT”, *Ínsula. Revista de Letras y Ciencias Humanas*, 822 (junio de 2015), pp. 1-30.
- ALLÉS-TORRENT, Susanna, “Sobre la complejidad de los datos en Humanidades digitales, o cómo traducir las ideas a datos”, *Revista de Humanidades Digitales*, 4 (2019), pp. 1-28, <<https://doi.org/10.5944/rhd.vol.4.2019.24679>>.
- ALLÉS-TORRENT, Susanna, “Crítica textual y edición digital o ¿dónde está la crítica en las ediciones digitales?”, *Studia aurea: revista de literatura española y teoría literaria del Renacimiento y Siglo de Oro*, XIV (2020), pp. 63-98, <<https://doi.org/10.5565/rev/studiaaurea.395>>.
- CZEPIEL, Maria, «Two Newly Discovered Poems by Garcilaso de la Vega», *Bulletin of Spanish Studies*, 99, 5 (2022), pp. 741-776.
- FOSALBA, Eugenia y DE LA TORRE, Gáldrick ed., *La égloga renacentista en el reino de Nápoles*, *Bulletin Hispanique*, Presses Universitaires de Bordeaux, 2017.
- FOSALBA, Eugenia y DE LA TORRE, Gáldrick, *Contexto latino y vulgar de Garcilaso en Nápoles: redes de relaciones de humanistas y poetas (manuscritos, cartas, academias)*, Peter Lang Verlag, 2018.
- FOSALBA, Eugenia, “Descripción del Ms. XIII AA 63 de la BN de Nápoles, transmisor de dos odas neolatinas de Garcilaso”, *Contexto latino y vulgar de Garcilaso en Nápoles: redes de relaciones de humanistas y poetas (manuscritos, cartas, academias)*, Peter Lang Verlag, 2018, pp. 297-321.
- FOSALBA, Eugenia, y DE LA TORRE, Gáldrick, “*Di qui Spagna et Italia han mostro chiaro l'onor*”. *Estudios dedicados a Tobia R. Toscano sobre Nápoles en tiempos de Garcilaso*, Studia Aurea Monográfica, Publicaciones Universitat Autònoma de Barcelona, 2019.
- Fosalba, Eugenia, ed. «Clasicismo horaciano en tiempos de Garcilaso de la Vega», *Bulletin Hispanique*, 125-1, 2023.
- REHBEIN, Malte, y Christiane FRITZE, “Hands-On Teaching Digital Humanities: A Didactic Analysis of a Summer School Course on Digital Editing”, en *Digital Humanities Pedagogy: Practices, Principles and Politics*, ed. Brett D. Hirsch, Cambridge, Open Book Publishers, 2015, pp. 47-78, <<http://books.openedition.org/obp/1617>>.
- SAHLE, Patrick, “What Is a Scholarly Digital Edition”, en *Digital Scholarly Editing: Theories and Practices*, ed. Matthew James Discroll y Elena Pierazzo, Open Book Publishers, 2016, <<https://doi.org/10.11647/OBP.0095>>.

Procesamiento del lenguaje natural y fijación del texto. Experiencias en torno a la constitución de un corpus diacrónico de sonetos

Helena Bermúdez Sabel

JinnTec
helena.bermudez@jinntec.de

Clara Isabel Martínez Cantón

UNED
cimartinez@flog.uned.es

Pablo Ruiz Fabo

Université de Strasbourg
ruizfabo@unistra.fr

Resumen

Esta contribución surge en el contexto de desarrollo del corpus de sonetos DISCO (*Diachronic Spanish Sonnet Corpus*), un corpus de 4530 sonetos en español compuestos entre el siglo XVI y el XX por autores de diversas procedencias (Europa, Latinoamérica y Filipinas). Este recurso contiene las anotaciones de diferentes fenómenos de versificación que han sido obtenidas a partir de técnicas del procesamiento del lenguaje natural (PLN). En este artículo presentamos cómo los resultados de la anotación automática pueden ser utilizados para detectar problemas de transmisión textual. Uno de los objetivos de esta contribución es el de proporcionar claves sobre posibles flujos de trabajo que, ayudándose de herramientas de PLN, permitan detectar posibles errores textuales, centrando así los esfuerzos de revisión manual en pasajes concretos.

Palabras clave

corpus; soneto; procesamiento del lenguaje natural; anotación; edición digital; versificación.

Abstract

Natural Language Processing and Text Curation. Experiences within the Development of a Diachronic Sonnet Corpus.

We present work carried out within the development of DISCO, the *Diachronic Spanish Sonnet Corpus project*, which consists of 4,530 sonnets in Spanish from Europe, Latin America and the Philippines, including texts from the 15th to the 20th centuries. The resource offers versification annotations obtained automatically through tools based on Natural Language Processing (NLP). In this article, we present how automatic annotation results can be exploited to detect textual transmission errors. Drawing on our experience with DISCO, we present observations towards the creation of workflows assisted by NLP-based tools, which can help detect possible textual errors, thus allowing us to focus on specific passages for our manual correction effort.

Keywords

Corpus; sonnet; natural language processing; annotation, scholarly editing, versification.

Introducción

El propósito de trabajar con corpus que reúnan una colección numerosa de textos literarios, que permitan la comparación de sus características genéricas, métricas, estilísticas, etc., se había visto condicionada, hasta la era digital, por la capacidad limitada del ser humano para recopilar, clasificar y analizar datos. En otras disciplinas, como la lingüística de corpus, este enfoque metodológico tuvo más presencia. Sin embargo, también en los estudios literarios encontramos previamente a la utilización de bases de datos o internet, trabajos de investigación que llevan a cabo cuantificaciones sobre determinadas características de un corpus. En el área de la métrica poética cabe nombrar, por ejemplo, repertorios que recogían los esquemas métrico-rimáticos de un corpus, y que fueron creados mediante un análisis manual, como *Die nicht lyrischen Strophenformen des Altfranzösischen* de Gotthold Naetebus (1891), repertorio métrico que reunía con un método riguroso la poesía narrativa francesa altomedieval, conservado en un comienzo únicamente en papel.

La edición de textos es un campo que se ha visto enriquecido por la utilización de métodos digitales de muy diversas maneras. En este artículo queremos exponer cómo la recopilación y anotación con herramientas digitales del

corpus DISCO (*Diachronic Spanish Sonnet Corpus*)¹ ha servido para mejorar la edición y curación de este corpus, pues ha permitido encontrar y corregir errores en la transmisión textual. Presentaremos, así, cómo ciertos indicios que percibimos gracias a la anotación automática pueden ser indicativos para la detección de pasajes problemáticos que pueden (o deben), después, ser co-tejados manualmente.

DISCO es un corpus en continuo crecimiento,² que recoge en el momento presente 4530 sonetos en español de distintas épocas y lugares. Recoge poemas de 1215 autores y autoras procedentes de 22 países distintos de Europa, Latinoamérica y Filipinas. Cronológicamente se recopilan sonetos desde los inicios de esta forma poética (los primeros ensayos del Marqués de Santillana) hasta sonetos modernistas con versos alejandrinos o incluso hexadecasílabos.

Los sonetos recogidos en DISCO se extrajeron en la mayoría de los casos de la Biblioteca Virtual Miguel de Cervantes, de distintas colecciones de sonetos (García González 2006a; 2006b), u otros recursos, como el Portal de Literatura Hispanofilipina. Algunos textos del siglo XVIII proceden de Wikisource u otras fuentes, y para la elaboración del subcorpus modernista el equipo de investigación ha realizado la digitalización de diferentes ediciones impresas. Estas diversas fuentes han sido homogeneizadas y estandarizadas gracias a su transformación en XML, siguiendo las directrices de la Text Encoding Initiative (TEI).³ Otro valor añadido de DISCO, sin embargo, consiste en enriquecer estos textos, añadiéndoles, de manera sistemática, distinta información. Por una parte, se han añadido metadatos para los autores, como las fechas de nacimiento, muerte, procedencia y el identificador de VIAF (Virtual International Authority File). El Identificador VIAF⁴ (Fichero de Autoridades Virtual Internacional) agrupa en un solo registro los diferentes registros de autoridad provenientes de diversas instituciones internacionales, en su mayoría bibliotecas nacionales. El identificador VIAF permite la identificación y referenciación consistente de los y las poetas del corpus, por lo que es relevante incluirlo.

Además, y lo que, creemos, hace más interesante al corpus, DISCO incluye anotaciones literarias, creadas con distintas herramientas digitales (ver, para más detalle Ruiz Fabo *et al.* 2021). El corpus está analizado métricamente, recogien-

1. <https://github.com/pruizf/disco> (último acceso el 26 de septiembre de 2023).

2. Ver <https://github.com/pruizf/disco/#versions> (último acceso el 26 de septiembre de 2023).

3. La Iniciativa de Codificación de Texto (TEI, por sus siglas en inglés) es un consorcio que desarrolla y mantiene de manera colectiva unas directrices para la representación de textos en formato digital. Este sistema de representación es el lenguaje de marcado, actualmente serializado como un vocabulario XML comúnmente denominado XML-TEI, que cuenta con una guía de uso mantenida por el consorcio. Para obtener más información sobre la Text Encoding Initiative (TEI), puede visitar su sitio web oficial: <https://tei-c.org/>.

4. <https://viaf.org/>

do el número de sílabas de cada verso, su acentuación, su rima e incluso la existencia o no de encabalgamiento.⁵

Este corpus, además, puede consultarse a partir de una interfaz que permite la búsqueda y exploración sencilla de los sonetos y la visualización con gráficos dinámicos de sus características métrico-literarias, DISCO⁶ (Bermúdez Sabel *et al.*, 2022).

Buscando la reutilización y la posible investigación con el corpus, DISCO se ofrece tanto en texto simple como en XML-TEI, este último enriquecido con RDFa, un formato de datos enlazados,⁷ utilizando por tanto estándares con gran proyección en las humanidades.

Así, este corpus contribuye al desarrollo de recursos digitales para el estudio de la poesía en español, especialmente, desde una perspectiva comparativa, diacrónica y polisistémica.

La anotación automática va seguida de una labor de revisión y edición manual para la que ciertos aspectos de la anotación son clave, como indicamos en un comienzo. Es interesante recalcar que estos aspectos, además, no son los mismos para el Siglo de Oro que para el Modernismo, porque las tendencias literarias y la métrica también van evolucionando.

Presentaremos aquí el proceso y algunas muestras del trabajo de edición realizado con el subcorpus del Siglo de Oro, que recoge un total de 477 poetas que son autores de 1088 sonetos. Es interesante señalar la existencia de otro corpus distinto de sonetos de Siglo de Oro, el *Corpus de Sonetos del Siglo de Oro* (Navarro-Colorado *et al.*, 2016), que se solapa solo en una pequeña parte con DISCO, pues 17 de los autores se hallan representados en ambos corpus y encontramos 333 sonetos en común.

Retos

Los desafíos encontrados para la edición de este corpus de sonetos han sido de diversa naturaleza.

Hay algunos retos que pueden parecer comunes a la recopilación de un corpus, sea este digital o no, como todo el flujo de trabajo relacionado con la

5. Para estas anotaciones se han utilizado herramientas como Jumper (Marco Remón y Gonzalo 2021), el sistema de ADSO (Navarro-Colorado 2017) y Rantanplan (de la Rosa *et al.* 2020) para la escansión y la acentuación métricas, Rhyme Tagger (Plecháč 2018) para la rima y ANJA (Martínez Cantón *et al.* 2021; Ruiz *et al.* 2017) para el encabalgamiento. Las anotaciones métricas han sido corregidas manualmente en los sonetos modernistas, mientras que las de rima han corrido la misma suerte en el subcorpus del Siglo de Oro.

6. <https://prf1.org/disco/> (último acceso el 26 de septiembre de 2023).

7. El corpus ha sido utilizado para diversos estudios de humanidades digitales, como se recoge en <https://github.com/pruizf/disco/#resources-reusing-disco> (último acceso el 26 de septiembre de 2023).

selección de los elementos, en este caso los sonetos, del corpus, y la información que sobre ellos se proporciona. Sin embargo, como es bien sabido, el medio condiciona el contenido, y ofrece posibilidades y limitaciones. En nuestro caso, condicionó la selección de un solo tipo de composición poética, el soneto. Esta elección ha sido en parte debida a la incontestable relevancia y tradición de esta forma composicional en la literatura española. No es únicamente una forma poética, sino que puede ser considerada un subgénero lírico (Romero Luque 2017; Spang 1993; Zavala 1995). No obstante, las razones que nos guiaron a recoger un corpus solo de sonetos tienen que ver también con su tratamiento digital. Se trata de una forma manejable desde el punto de vista computacional, ya que obedece a restricciones claras. La variabilidad se mantiene dentro de unos límites, lo que facilita la comparación de rasgos métricos entre poemas, o entre distintos subcorpus de poemas.

Además, cuando este corpus comienza a gestarse, las herramientas de escanión y asignación del acento métrico estaban más avanzadas para el endecasílabo, mientras que para otras formas métricas la precisión era menor o los resultados eran poco significativos. No en vano, usando métodos computacionales, el soneto había sido estudiado ya por distintos investigadores (Navarro-Colorado 2015; Navarro-Colorado, Lafoz, y Sánchez 2016; Navarro-Colorado 2017; Agirrezabal, Alegria, y Hulden 2017). Por estas razones, un nuevo corpus de sonetos enriquecía el panorama ya existente y abría un diálogo con trabajos anteriores en los estudios literarios tradicionales, en el desarrollo de corpus digitales y en el análisis computacional de la poesía.

La elección de las fuentes de las que se tomarán los sonetos supuso también un reto, ya que, por una parte, debíamos aprovechar el material ya digitalizado disponible en bibliotecas digitales, pero éramos conscientes de que debía construirse un corpus que siguiera unos determinados criterios de selección con el fin de ofrecer a la comunidad científica un recurso de gran utilidad. Así, y dado que existía ya el corpus del Siglo de Oro (siglos xv-xvii) de Navarro-Colorado (2015) centrado en autores canónicos, DISCO se ocupa de recopilar sonetos de todo tipo de autores, incluyendo autores menores, que completen (y complementen) así otro tipo de acercamientos.

Con la idea de recoger una muestra de cómo se ha cultivado el soneto en español desde el siglo xv hasta el Modernismo en el siglo xx, no solo desde los autores canonizados, sino desde una perspectiva amplia, se recogieron sonetos de todas las épocas, potenciando así la dimensión diacrónica. Además, las dos últimas versiones de DISCO,⁸ de 2021 y 2023 se recogen poemas de autores filipinos y se le da más cabida a escritores latinoamericanos, tratando de ampliar también la procedencia de los poetas del corpus.

8. Consultar aquí: <https://github.com/pruizf/disco/#versions> (último acceso el 26 de septiembre de 2023).

Otro reto que se nos planteó fue el de incrementar la usabilidad del corpus y, de manera particular, hacer visible la existencia de poetas mujeres, y la procedencia de los autores. Para ello, a través de la interfaz de exploración del corpus DISCOOver se facilitó la consulta y navegación de corpus a través de dos funciones de búsqueda principales: búsqueda por autor y búsqueda facetada basada en metadatos de autor. El buscador dispone de filtros para periodos, género (femenino o masculino) y origen del autor o autora (Latinoamérica, Asia o Europa).

Además, con la intención de facilitar la selección y reutilización de subcorpus derivados de DISCO, DISCOOver permite seleccionar los sonetos o autores deseados y etiquetar el subcorpus creado. Este nombre se utilizará para identificar los resultados del subcorpus en los gráficos que la herramienta genera. Para los poemas seleccionados por el usuario, se crean gráficos dinámicos que muestran diferentes características métrico-rimáticas de la selección. Pueden combinarse en un mismo gráfico los resultados de diferentes selecciones con el fin de facilitar la comparación cuantitativa. Es posible además acceder además al texto de los propios sonetos en los que se ofrece la anotación de diferentes elementos de versificación (patrón acentual de cada verso, esquema de rima y encabalgamiento). Por otra parte, la creación de nuevas herramientas de anotación automática ha dado nuevas posibilidades a este corpus. Así, si antes nos habíamos ceñido al soneto hasta el siglo XIX, pues son casi sin excepción endecasilábicos, las versiones nuevas de DISCO incluyen sonetos modernistas, con sus distintas medidas, versos compuestos, etc. Su etiquetado automático ha sido posible gracias a la existencia de nuevas herramientas de escansión, como Jumper (Marco Remón y Gonzalo 2021) o Rantanplan (de la Rosa *et al.* 2020). Un tratamiento eficaz del verso compuesto ha sido posible gracias a Jumper, cuyos resultados con este tipo de versos (como evaluado en Marco Remón y Gonzalo 2021) superan los de otras herramientas.

La detección de la rima se añadió, asimismo, únicamente a partir de la versión 3 del corpus, publicada en octubre de 2019, con la adaptación de la herramienta Rhyme Tagger (Plecháč 2018) para el español.

La edición digital de textos cuenta ya con directrices específicas, por lo que DISCO, desde un primer momento ha seguido de cerca los criterios de RIDE para Colecciones de Textos Digitales (Henny-Krahmer y Neuber 2017).

La creación de un flujo de trabajo y unos criterios homogéneos para la edición y limpieza de los textos será, precisamente, el reto del que hablaremos en este artículo.

Metodología

Las anotaciones prosódico-rimáticas del corpus DISCO han sido realizadas a partir de diferentes herramientas de procesamiento del lenguaje natural. En el

caso particular de la rima, se ha utilizado el sistema multilingüe *Rhyme Tagger* (Plecháč 2018).

Uno de los mecanismos implementados en la elaboración del corpus DISCO para mejorar la calidad de este recurso ha sido el de revisar manualmente los versos no rimados. La rima se presenta como un fenómeno especialmente apropiado para focalizar los esfuerzos de revisión manual por diferentes motivos relacionados tanto con la historia de la versificación como con el procesamiento automático. Por una parte, dependiendo del período, la presencia de versos no rimados en el soneto en español puede considerarse una rareza (Pardo 2014). Un estudio sobre cómo se componía un soneto en el Siglo de Oro español señala precisamente la rima como piedra de toque sobre la que se construía el edificio poemático: “cuando uno se ponía a hacer un soneto, lo primero que hacía era apuntar en el margen las palabras rimantes. Esto no debe sorprender: es obvio que lo que rige todo soneto es la rima” (Dadson 1998: 512). Resulta llamativo, de hecho, que Kurt Spang (1993: 99-103), en el espacio que dedica al soneto como subgénero lírico, haga alusión a las posibles muchas variedades de la rima y que no se considere la posibilidad de los sonetos sin rima.

Por otra parte, el rendimiento de la anotación automática de la rima es muy elevado, con una precisión de 0.98 y una exhaustividad de 1. Esto quiere decir que la herramienta, en principio, detecta todos los casos de rima y que solo en 2 de cada 100 casos anota como rima algo que en realidad no lo es.

Teniendo en cuenta lo anterior, la presencia de un verso no rimado en el corpus de sonetos del Siglo de Oro puede relacionarse con tres posibles escenarios: 1) innovación formal, 2) problema de transmisión textual de la obra o 3) error generado durante el proceso de conformación del corpus. En otros períodos del corpus DISCO, en los que se ha generado una versión electrónica del texto a partir de fuentes impresas, tendríamos un caso específico del escenario 3 que es muy habitual: errores generados durante el proceso de digitalización (OCR). Sin embargo, la fuente del subcorpus del Siglo de Oro es nativamente digital (García González 2006). Es posible, no obstante, que nos encontremos con errores textuales en esta fuente que en su origen provengan de errores de OCR y que han pasado desapercibidos por el editor: en todo caso, en nuestro contexto específico tienen que ser categorizados como problemas de transmisión del texto, pues estamos utilizando una fuente digital curada.

Esta contribución explora cómo podemos utilizar los resultados de la anotación automática relativos a la rima para detectar posibles problemas en la transmisión textual de una obra. La búsqueda de estos potenciales problemas se centra en la presencia de versos no rimados que serán categorizados atendiendo a los tres posibles escenarios previamente descritos.

Algunos de los casos encontrados presentan una lectura que de manera inmediata puede ser corregida. Véase como ejemplo los siguientes versos de González de Andrade (en negrita, palabra final del verso no rimado):

Des que en lechos de zafir **reposan**
 y que por sendas de cristal caminas,
 derramando tus urnas cristalinas
 en favor de las playas arenosas;
 (Paulo Gonzálvez de Andrade, *Soneto*, vv. 1-4)

Desde un punto de vista semántico y formal, la última palabra del primer verso tiene que corregirse en “reposas”. La presencia de “reposan” es muy probablemente resultado del proceso de conformación de la edición digital de la Biblioteca Virtual Miguel de Cervantes, ya que esta es la única edición que hemos encontrado con la lectura “reposan” (y en los recursos de ella derivados).

Como decíamos, la búsqueda automática de versos no rimados nos ha proporcionado ejemplos que responden a un problema de fijación de texto y no a simples errores de digitalización. Este es el caso de uno de los versos no rimados encontrados en la obra de Esteban Manuel de Villegas en la que uno de sus sonetos presenta la siguiente lectura:

Quién me dijera, Clori, que algún día
 te pudiera olvidar tan fácilmente,
 mientras soltero crin hizo en tu frente
 con hilos de oro lazos de **rabia**.
 (Esteban Manuel de Villegas, *Eróticas*, Soneto II, vv. 1-4)

Ediciones recientes de esta antología de Villegas (p. ej. Villegas 2010) reproducen también la lectura “rabia”. Sin embargo, en las primeras ediciones de la obra de Villegas podemos encontrar una lectura mucho más adecuada, tanto desde el punto de vista semántico como formal: “taujiá” como palabra rima (véase Villegas 1618; 1774; 1797). Taujiá aparece hoy en día en el DRAE como palabra en desuso para “ataujía”, que significa, a su vez, damasquinado, es decir, una técnica artesanal de ornamentación que consiste en incrustar hilos de oro o plata en superficies metálicas, como acero o hierro, creando intrincados diseños decorativos. Como vemos, este término, además de encajar métricamente mucho mejor, tiene sentido en su contexto y resulta mucho más lírico.

Se reproduce a continuación una transcripción semi-diplomática de estos versos en la edición príncipe (Villegas 1618, 80):

Quien me dixera, Clori, que algun dia
 te pudiera olvidar tan facilmente,
 mientras foltero crin hiço en tu frente
 con hilos de oro laços de tauxia,
 (Esteban Manuel de Villegas, *Eróticas*, Soneto II, vv. 1-4)

La transmisión de la obra de Villegas parece mostrar una cierta complejidad, pues encontramos en los tercetos de uno de sus sonetos dos versos no rimados que no podemos corregir de manera inmediata:

Quando mostrabas de azucena, **fruta**,
 la tez bruñida, o como sin cuidado
 de mí solicitabas tus placeres,
 déjame pues, que si te quise **hermosa**,
 ya no es posible, puesto que has llegado
 a tiempo, que a ti misma no te quieres.

(Esteban Manuel de Villegas, *Eróticas*, Soneto X, vv. 9-14)

Una vez más, en una de las ediciones contemporáneas a Villegas (Villegas 1618, 82) encontramos una lectura con rima:

Quando mostrabas de açuçena, i rofa
 la tez bruñida; ó como fin cuidado
 de mi folicitabas tus placeres,
 Dejame pues, que fi te quife hermofa,
 ya no es posible, puesto que has llegado
 a tiempo, que a ti misma no te quieres.

(Esteban Manuel de Villegas, *Eróticas*, Soneto X, vv. 9-14)

En este caso, la hipótesis que manejamos para explicar cómo la lectura “azucena, y rosa” ha pasado a “azucena, fruta,” pasa por considerar que se trata de un error generado durante la creación de un recurso electrónico usado como fuente, causado por un error de OCR, y que los editores posteriores no han detectado. Es posible que el caso anteriormente comentado, la sustitución de *taujiá* por *rabia*, responda a la misma casuística.

Discusión

El trabajo de revisión de los versos no rimados, detectados automáticamente, ha sido una estrategia eficaz para mejorar la calidad de los textos contenidos en el corpus DISCO. Cabe destacar, no obstante, que la mayor parte de los versos no rimados detectados en el corpus respondía a errores de la herramienta de anotación, generalmente relacionados con rimas imperfectas (p. ej. *serenas* / *cantine-las*) o con grafías homófonas (p. ej. *donna* / *razona*, *divino* / *digno*). Por lo tanto, este proceso de revisión también nos ha llevado a implementar correcciones en la anotación rimática.

Si bien es cierto que los flujos de trabajo automatizados requieren de la intervención humana para mejorar los resultados de la automatización, en esta contribución argumentamos que las herramientas de procesamiento del lenguaje natural también pueden ser utilizadas para detectar algunos de los errores humanos. Esa misma sugerencia es hecha por los autores de dos herramientas de escansión automática para español (Rantanplan y Jumper) cuando señalan:

Remarkably, using both systems [Rantanplan and Jumper] on the same corpora has proved to be particularly useful in our experimentation: when both systems make the same mistake, it is a reliable signal that the manual annotation might be erroneous. Therefore, used in conjunction they can be a useful tool for corpus clean-up (Marco *et al.* 2021)

El corpus DISCO nace ante la necesidad de crear una gran colección de poemas en español (concretamente de sonetos) para poder responder a nuevas preguntas de investigación utilizando las metodologías habilitadas por la computación. Nos atrevemos a afirmar que entre las tareas necesarias para poder investigar en disciplinas de Humanidades Digitales más ingratas destaca la creación de los corpus: es un trabajo arduo y cronofágico, y que muchos organismos de evaluación académica en Humanidades ni siquiera consideran como producción científica.

El desarrollo de grandes colecciones de recursos electrónicos está generalmente vinculado a procesos de automatización que hagan esta tarea abarcable en el menor tiempo posible. La presencia de errores en el texto o en la anotación es prácticamente inevitable, de ahí que sea necesario establecer procesos de control de calidad en cada una de las etapas del flujo de trabajo (especialmente en las automatizadas).

Si bien es evidente que la calidad de los resultados de investigación adquiridos dependerá de la calidad científica del recurso utilizado, también es cierto que las metodologías computacionales nos permiten establecer márgenes de error aceptables: la presencia de errores no es incompatible con la calidad científica ni con la consecución de resultados pertinentes (véase, p. ej., Franzini *et al.* 2018). A la hora de desarrollar un recurso electrónico, resulta fundamental que el equipo de investigación de Humanidades defina cuáles serán los tests que permitan evaluar la calidad de los resultados de cada uno de los procesos de automatización y establecer un margen de error aceptable. A la hora de explotar dicho recurso, se necesitará hacer una reevaluación del margen de error atendiendo a los objetivos de la investigación específica que se va a desarrollar.

La Biblioteca Virtual Miguel de Cervantes es un recurso de gran valor para los estudios literarios en español. Aunque en esta contribución hayamos reflejado algunos de los problemas textuales que presenta una de las ediciones disponibles en este portal, consideramos que esta edición (dada su extensión) se encuentra dentro de los parámetros del margen de error aceptable.

En la elaboración del corpus DISCO hemos establecido un sistema de versiones documentado que permite una mejora continua (pero organizada) de este recurso. Esta estrategia ha permitido que una versión inicial del corpus estuviese disponible para poder ser utilizada por la comunidad científica ya en 2017. Seis años más tarde, hemos publicado la versión 5.0 del corpus (Ruiz Fabo, Bermúdez Sabel, y Martínez Cantón 2023). Este sistema de versiones permite que las y los usuarios de este recurso puedan citar la versión concreta sobre la que se basa

su trabajo y que el estado del corpus en ese momento siempre esté disponible. Al mismo tiempo, también pueden comprobar qué cambios ha sufrido el corpus en versiones posteriores y evaluar las nuevas posibilidades de investigación que les ofrece cada una de las actualizaciones.

Conclusiones

La presente contribución es un ejemplo de cómo la evaluación de los resultados de anotaciones automáticas (usando una o varias herramientas para la comparación) se puede utilizar como estrategia para mejorar el contenido textual de grandes corpus. En este caso, la anotación rimática automática nos ha guiado en tareas de fijación de texto limitadas a la palabra rima. No obstante, las correcciones manuales de anotaciones automáticas pueden explotarse de distintas maneras, pues cabe la posibilidad de que sirvan para mejorar herramientas de anotación existentes, para mejorar datos de entrenamiento y también para la creación de modelos de corrección automática.

Esta contribución y el trabajo de revisión textual llevado a cabo dentro del corpus DISCO es igualmente útil para otros recursos derivados de la Biblioteca Virtual Miguel de Cervantes como ADSO (Navarro-Colorado, Ribes-Lafoz y Sánchez 2016).

Las ediciones automatizadas son imperfectas, pero valiosas. En el análisis donde se busca evaluar estadísticamente una tendencia (lectura distante, lingüística de corpus), la existencia de cierta proporción de errores no es un obstáculo insalvable, ya que las tendencias se pueden detectar dentro de ciertos márgenes de error.

En este sentido, una edición automatizada de un corpus amplio puede ayudar a detectar tanto las tendencias generales como los valores atípicos. Así, la detección, gracias a la anotación automática, de pasajes en los que encontramos características no esperadas (número de versos, de sílabas, rimas inesperadas) puede ayudar también a focalizar nuestros esfuerzos críticos e interpretativos en las partes más innovadoras, y llegar a una mayor comprensión de la experimentación poética, con el objetivo de acometer estudios y ediciones críticas.

Como ampliación y generalización de este trabajo inicial, planeamos utilizar la anotación métrica para buscar irregularidades (o “rarezas”) en la escansión que nos permitan detectar errores de transmisión textual, así como errores generados durante la creación del corpus (mucho más frecuentes). El objetivo es precisamente el de eliminar todas las irregularidades causadas por errores, para poder centrarse en el estudio de las innovaciones o particularidades métrico-rimáticas.

Bibliografía

- AGIRREZABAL, Manex, Iñaki ALEGRIA, y Mans HULDEN, “A Comparison of Feature-Based and Neural Scansion of Poetry”, *RANLP 2017 - Recent Advances in Natural Language Processing Meet Deep Learning*, noviembre, 2017, pp. 18-23. <https://doi.org/10.26615/978-954-452-049-6_003>.
- BERMÚDEZ SABEL, Helena, Pablo RUIZ FABO, et Clara MARTÍNEZ CANTÓN, “DISCOVERing Spanish Sonnets: A circular reading experience”, en *Computational Stylistics in Poetry, Prose and Drama*, ed. Anne-Sophie Bories, Petr Plecháč, y Pablo Ruiz. De Gruyter. <<https://doi.org/10.1515/9783110781502-004>>
- DADSON, Trevor J. 1998. “Cómo se hacía un soneto en el siglo de oro: el caso de ‘amor, la red de amor digo que es hecha’”, en *Siglo de Oro. Actas del IV Congreso Internacional de la AISO*, Alcalá de Henares, 1996, pp. 2: 509-524.
- FRANZINI, Greta, Mike KESTEMONT, Gabriela ROTARI, Melina JANDER, Jeremi K. OCHAB, Emily FRANZINI, Joanna BYSZUK, y Jan RYBICKI, “Attributing Authorship in the Noisy Digitized Correspondence of Jacob and Wilhelm Grimm”, *Frontiers in Digital Humanities*, 5 (2018). <<https://www.frontiersin.org/articles/10.3389/fdigh.2018.00004>>.
- GARCÍA GONZÁLEZ, Ramón, ed., *Sonetos del siglo XIX*, Alicante, Biblioteca Virtual Miguel de Cervantes, 2006a, <<https://www.cervantesvirtual.com/nd/ark:/59851/bmc4q861>>.
- GARCÍA GONZÁLEZ, Ramón, ed., *Sonetos del siglo XV al XVII*, Alicante, Biblioteca Virtual Miguel de Cervantes, 2006b, <<https://www.cervantesvirtual.com/nd/ark:/59851/bmc2r439>>.
- HENNY-KRAHMER, Ulrike, y Frederike NEUBER, “Criteria for Reviewing Digital Text Collections, version 1.0”, *IDE. A Review Journal for Digital Editions and Resources*, 6 (2017), <<https://www.i-d-e.de/publikationen/weitereschriften/criteria-text-collections-version-1-0/>>.
- MARCO REMÓN, Guillermo, y Julio GONZALO, “Escansión automática de poesía española sin silabación”, *Procesamiento del Lenguaje Natural* 66 (2021), pp. 77-87, <<https://doi.org/10.26342/2021-66-6>>.
- MARTÍNEZ CANTÓN, Clara Isabel, Pablo RUIZ FABO, Elena GONZALEZ-BLANCO, Elena, y Thierry POIBEAU. 2021. “Automatic Enjambment Detection as a New Source of Evidence in Spanish Versification”, en *Plotting Poetry: On Mechanically Enhanced Reading*, ed. Anne-Sophie Bories, Gerald Purnelle, y Hugues Marchal, Liège, Presses Universitaires de Liège, pp. 93-112.
- NAETEBUS, Gotthold, *Die nicht-lyrischen strophformen des altfranzösischen*, Leipzig, Druck von J.B. Hirschfeld, 1891.
- NAVARRO-COLORADO, Borja, “A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects”, en *Proceedings of the Computational Linguistics for Literature Workshop at NAACL 2015*, Denver, U.S., Association for Computational Linguistics, 2015.
- NAVARRO-COLORADO, Borja, “A metrical scansion system for fixed-metre Spa-

- nish poetry”, *Digital Scholarship in the Humanities* 33.1 (2017), pp. 112-27. <<https://doi.org/10.1093/llc/fqx009>>.
- NAVARRO-COLORADO, Borja, María RIBES LAFOZ, y Noelia SÁNCHEZ, “Metrical Annotation of a Large Corpus of Spanish Sonnets: Representation, Scansion and Evaluation”, en *Proceedings of the Tenth International Conference on Language Resources and Evaluation, Portoroz, Slovenia*, Portorož, Slovenia, 2016, pp. 4630-34. <http://www.lrec-conf.org/proceedings/lrec2016/pdf/453_Paper.pdf>.
- PARDO, Arcadio, “De la diversidad del soneto”. *Rhythmica. Revista Española de Métrica Comparada*, 12 (2014), pp. 127-172, <<https://doi.org/10.5944/rhythmica.14247>>.
- PLECHÁČ, Petr, “A Collocation-Driven Method of Discovering Rhymes (in Czech, English, and French Poetry)”, en *Taming the Corpus: From Inflection and Lexis to Interpretation*, ed. Masako Fidler y Václav Cvrček, Cham, Springer International Publishing, 2018, pp. 79-95. <https://doi.org/10.1007/978-3-319-98017-1_5>.
- ROMERO LUQUE, Manuel, “El soneto modernista (Manuel Machado como paradigma)», *Rhythmica*, 15 (2017), pp. 113-45. <<https://doi.org/10.5944/rhythmica.21190>>.
- ROSA, Javier de la, Álvaro Pérez, Laura HERNÁNDEZ, Salvador Ros, y Elena GONZÁLEZ-BLANCO, “Rantanplan, fast and accurate syllabification and scansion of spanish poetry”, *Procesamiento del Lenguaje Natural*, 65 (2020), pp. 83-90.
- RUIZ FABO, Pablo, Helena BERMÚDEZ SABEL, Clara MARTÍNEZ CANTÓN, y Elena GONZÁLEZ-BLANCO, “The Diachronic Spanish Sonnet Corpus: TEI and Linked Open Data Encoding, Data Distribution, and Metrical Findings”, *Digital Scholarship in the Humanities*, 36 (Supplement_1) (2021), pp. i68-80, <<https://doi.org/10.1093/llc/fqaa035>>.
- RUIZ FABO, Pablo, Clara MARTÍNEZ CANTÓN, Thierry POIBEAU, y Elena GONZÁLEZ-BLANCO, “Enjambment Detection in a Large Diachronic Corpus of Spanish Sonnets”, en *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, 2017, pp. 27-32, <<http://www.aclweb.org/anthology/W17-2204>>.
- SPANG, Kurt, *Géneros literarios*, Madrid, Síntesis, 1993.
- VILLEGAS, Esteban Manuel de, *Las eróticas o amatorias*, Nájera, por Ivan de Mongastón, 1618.
- VILLEGAS, Esteban Manuel de, *Las Eróticas; y Traducción de Boecio*, Madrid, por Don Antonio de Sancha, 1774.
- VILLEGAS, Esteban Manuel de, *Las Eróticas y traducción de Boecio*, Segunda edición, Madrid, por Don Antonio Sancha, 1797.
- VILLEGAS, Esteban Manuel de, *Las eróticas o Amatorias*. ed. María Ángeles Díez Coronado y Emilio Magaña Orúe, Logroño, Instituto de Estudios Riojanos, 2010.

ZAVALA, Iris M, “El amor es una aventura en el mal: los sonetos de Sor Juana”, *Tropelias: Revista de Teoría de la Literatura y Literatura Comparada*, 5-6 (diciembre) (1995), pp. 443-52, <https://doi.org/10.26754/ojs_tropelias/tropelias.19955-65581>

Edición digital enriquecida: un modelo de anotación multinivel para poesía del Siglo de Oro

Borja Navarro Colorado

Universidad de Alicante
borja@dlsi.ua.es

Resumen

En este capítulo se presenta un modelo general para la anotación multinivel de corpora de texto literario. Por multinivel se hace referencia a la combinación, en un mismo corpus, de información de diferentes niveles de descripción lingüística o literaria, desde datos relacionados con palabras o sílabas, hasta cuestiones temáticas, textuales o pragmáticas. El objetivo final de un corpus de estas características es fijar un posible análisis literario, por lo que se considera como una edición digital enriquecida. Se defienden cuatro características que un corpus de texto literario debe cumplir: interoperabilidad, perspectivismo, unidad y claridad/sencillez. Se da cuenta de los principales problemas de formalización en un corpus multinivel de este tipo: la combinación de diferentes formalismos de representación y, en el caso de XML, el problema de un anidamiento incorrecto. Finalmente se propone un modelo para un corpus de poesía del Siglo de Oro.

Palabras clave

corpus multinivel; poesía; métrica; sintaxis; anotación; edición digital; XML; TEI.

Abstract

Enriched Digital Edition: a Multilevel Annotation Model for Golden-Age Spanish Poetry.

This paper presents a general model for the multilevel annotation of a literary corpus. Multilevel refers to the combination of information from different linguistic or literary levels in the same corpus: from word related data to thematic, textual or pragmatic questions. The objective is to fix a possible literary analysis. To be considered an enriched digital edition, an annotated corpus must meet four

characteristics: interoperability, perspectivism, unity and clarity/simplicity. The main formalization problems are discussed: the combination of different representation formalisms and, in the case of XML, the improper nesting. Finally, a model for a corpus of poetry from the Spanish Golden-Age is proposed.

Keywords

multilevel corpus; poetry; meter; syntax; annotation; digital edition; XML; TEI.

Procesamiento del lenguaje natural y los estudios literarios computacionales¹

Una de las áreas de los llamados estudios literarios computacionales (CLS por sus siglas en inglés)² es la aplicación de herramientas de procesamiento del lenguaje natural (PLN) e inteligencia artificial (IA) al estudio del texto literario. Esta aplicación presenta diversos problemas.

En primer lugar, las herramientas de PLN, salvo casos excepcionales, se crean para procesar un texto moderno estándar. El análisis automático del texto literario es mucho más complejo por dos motivos principalmente. Por un lado porque el texto literario suele ser texto antiguo, con variantes históricas del idioma (aunque esté modernizado) para las que las herramientas de PLN no están preparadas. A esto se le une la desautomatización propia del estilo literario, por la cual se exprime al máximo la capacidad expresiva del idioma y se buscan formas lingüísticas novedosas.

Por otro lado, las herramientas de PLN no suelen estar diseñadas para extraer información literaria. Éstas analizan aquellos aspectos lingüísticos que la comuni-

1. Trabajo parcialmente financiado por el Ministerio de Ciencia e Innovación a través del proyecto “CORTEX: Conscious Text Generation” (PID2021-123956OB-I00); MCIN/AEI/10.13039/501100011033/ y “FEDER Una manera de hacer Europa”; y por la Generalitat Valenciana (Conselleria d’Educació, Investigació, Cultura i Esport) a través del Proyecto: NL4DIS-MIS: Tecnologías del Lenguaje Natural para lidiar con la desinformación (CIPROM/2021/021).

2. Para una visión general de los métodos actuales en los estudios literarios computacionales, véase Schöch *et al.* eds. (2023).

dad de PLN considera relevantes, como categorías gramaticales, sintaxis, sentimientos o entidades nombradas, entre otras.³ Esta información solo resulta en parte de interés para los estudios literarios. Son pocas las herramientas de PLN capaces de extraer rasgos específicos del texto literario, como puede ser la métrica (Navarro Colorado 2018a, de la Rosa *et al.* 2020, Haider 2021), figuras retóricas y metáforas novedosas (Kesarwani *et al.* 2017), análisis de personajes, detección de eventos narrativos y otros aspectos narratológicos (Piper *et al.* 2021), etc.

Para poder aplicar, por tanto, análisis computacional al texto literario es necesario desarrollar herramientas específicas capaces, por un lado, de interpretar el texto literario en su complejidad histórica y expresiva, y por otro, que sean capaces de extraer información de interés para los estudios literarios.

Para desarrollar estas herramientas hay que aplicar las técnicas propias del PLN. El paradigma técnico dominante para el desarrollo de herramientas de PLN en los últimos 30 años ha sido el aprendizaje automático o *machine learning* (Jurafsky y Martin 2023),⁴ y en concreto el aprendizaje automático supervisado.⁵ A diferencia de los métodos tradicionales en los que se indica a la máquina cómo analizar un texto mediante reglas de programación, en el paradigma de aprendizaje automático supervisado es la propia máquina la que infiere cómo analizar textos a partir de muestras textuales anotadas por humanos. Estas muestras textuales son los corpora anotados, que como se puede comprobar, tienen un papel básico en PLN, bien sea como corpus de aprendizaje en los métodos supervisados, bien como corpus de evaluación en cualquier caso (Pustejovsky y Stubbs 2012, Ide y Pustejovsky (eds.) 2017).

Con esta situación, para los estudios literarios computacionales en general, y para el desarrollo de herramientas de PLN específicas para texto literario en particular, es necesario desarrollar corpora de textos literarios anotados con información tanto lingüística como propiamente literaria.

La anotación de corpus se puede considerar además un modelo de análisis lingüístico y literario en sí mismo. Anotar un corpus es una tarea de interpretación

3. Véase Jurafsky y Martin (2023), en especial los capítulos 17 a 28, centrados en el análisis computacional de estructuras lingüísticas.

4. En el PLN actual sin embargo, dominan las técnicas neuronales y el desarrollo de grandes modelos de lenguaje (large language models o LLM), caracterizados precisamente por no necesitar corpora anotados a mano. Sin embargo, tanto para adaptar esos grandes modelos de lenguaje a tareas concretas (como las que se requieren en CLS), el llamado fine-tuning, como para las técnicas de prompting basadas en ejemplos (few-shots prompting), los corpora anotados sí son necesarios.

5. Las técnicas de aprendizaje automático se agrupan en dos grandes familias: las técnicas supervisadas y las no supervisadas. Las supervisadas parten de un corpus anotado, mientras que las no supervisadas parten de un corpus sin ningún tipo de marca ni anotación. En este trabajo me centro en la aproximación supervisada, que es la que más relevancia tiene en los corpora anotados. En los estudios literarios también se han aplicado técnicas no supervisadas, como por ejemplo *Topic Modeling* (Navarro Colorado 2018b). Sobre la aplicación de aprendizaje automático a los estudios literarios computacionales, véase Hatzel *et al.* (2023).

y estudio de la lengua literaria a partir del análisis de muestras textuales. Entre otros aspectos requiere definir un modelo teórico general, especificar qué información se va a marcar, formalizar esa información mediante un lenguaje de marcas, determinar el proceso de anotación, detectar casos complejos (normalmente problemas de ambigüedad) y determinar heurísticas para su correcta interpretación y anotación, entre otros aspectos (Ide y Pustejovsky (eds.) 2017). Aplicado a un texto literario, es una tarea hermenéutica propia de los estudios literarios (Meister 2022).

La anotación de corpus como paradigma de análisis lingüístico-literario no está exenta de problemas. Sin entrar en detalles, comentaré dos aspectos. Por un lado, la anotación tiene sus límites porque no se puede asumir que se pueda anotar todo tipo de información lingüístico-literaria. Por otro, toda formalización supone siempre un proceso de simplificación, y por tanto supone cierta pérdida en la precisión de análisis; en especial los casos más complejos (aquellos que se desvían de la norma general), que se obvian o se simplifican demasiado para ajustarlos al modelo general (McShane y Nirenburg 2001: 52-55). Como principal ventaja, la anotación del corpus permite automatizar el análisis literario y, así, realizar análisis a gran escala de amplios corpora.

Desde la óptica de los estudios literarios, un corpus anotado es lo que vamos a denominar una “edición digital enriquecida”, en tanto que edición digital de una o más obras en las que no solo se ha fijado el texto, sino que también se ha establecido una posible interpretación de las obras a uno o varios niveles de análisis: métrica, sintáctica, semántica, temática, etc. Y se ha realizado además con ánimo de ser análisis de referencia para futuras investigaciones del fenómeno literario tratado.

En este trabajo presentamos un modelo general para la anotación multinivel de corpora de texto literario. Por multinivel me refiero a un modelo de anotación que permita combinar información de diferentes niveles de descripción lingüística o literaria, de ahí el nombre de “multinivel”: desde datos relacionados con palabras o sílabas, hasta cuestiones temáticas, textuales o pragmáticas. En la siguiente sección se expondrán los requisitos que considero debe tener un corpus de texto literario anotado para considerarse edición digital enriquecida; luego comentaré los principales problemas de formalización que presenta el desarrollo de un corpus multinivel, y finalmente se propondrá un modelo general para un corpus de poesía del Siglo de Oro.

Requisitos generales para una edición digital enriquecida

Desarrollar un corpus anotado es una tarea compleja que requiere tiempo y esfuerzo. Se debe abordar siempre con la intención de conseguir una anotación útil y válida durante mucho tiempo. Así como en una edición crítica se busca fijar un texto de referencia para su lectura y análisis posterior por parte de la comunidad científica, una edición enriquecida busca marcar en el texto un aná-

lisis de referencia que sea útil para los estudios literarios durante los siguientes años. No me refiero con esto a establecer una única interpretación del texto literario (más sobre este punto luego), sino a establecer un análisis que sirva de referencia para otros análisis, bien como soporte para investigaciones relacionadas, bien para proponer análisis alternativos, o bien como corpus de aprendizaje para sistemas automáticos.

Para conseguir, por tanto, que un corpus anotado como el aquí planteado sea útil, fiable y válido para los estudios literarios, de manera que pueda ser considerado una edición digital enriquecida, debe cumplir cuatro requisitos básicos:

- interoperabilidad,
- perspectivismo,
- unidad
- sencillez y claridad.

A continuación comentaré cada uno de ellos.

Interoperabilidad

El concepto de interoperabilidad proviene de los sistemas de información y su capacidad para utilizarse de forma conjunta, compartir datos e intercambiar información.⁶ Así, entendiendo un corpus anotado como un sistema de información, debe estar diseñado de tal manera que la información contenida (la anotación) pueda ser compartida y reutilizada por otros corpora y/o por sistemas automáticos y herramientas de PLN.

Este rasgo recae sobre todo en la formalización. El formato de anotación debe ser lo más compatible posible, incluyendo aquí tanto el lenguaje de marcado utilizado, como la sistematización de la información realizada y la representación de ésta mediante etiquetas. Esto asegura la reutilización de la información anotada, su complementariedad con otros corpora y, en general, el intercambio de información con diferentes sistemas automáticos.

Para ello es necesario utilizar lenguajes y modelos de marcado estándar. En humanidades digitales y estudios literarios computacionales, el lenguaje de marcas más utilizado es el eXtensible Markup Language o XML usado según las recomendaciones de la *Text Encoding Initiative* o TEI (TEI Consortium 2023). Sin embargo, no siempre XML-TEI es la mejor opción. En la siguiente sección se expondrán sus límites y algunos problemas que pueden surgir al anotar un corpus multinivel con este formato estándar.

6. Véase cómo define interoperabilidad el *Diccionario panhispánico del español jurídico*: <https://dpej.rae.es/lema/interoperabilidad> (19 de septiembre de 2023).

Perspectivismo

En el área del PLN, la calidad de un corpus de aprendizaje se determina según la consistencia de la anotación (Pustejovsky y Stubbs 2012, Ide y Pustejovsky 2017): en qué medida, un mismo fenómeno lingüístico es anotado de la misma manera por dos personas diferentes. Se busca el máximo consenso en la anotación del corpus para que la máquina, al final, aprenda aquello en lo que los humanos están de acuerdo. El desacuerdo entre los anotadores se penaliza, pues se considera que hay un error en la anotación (bien sea en la formalización, en la sistematización, en las etiquetas, o en cualquier otro aspecto).

Este planteamiento es apropiado en la anotación lingüística de corpora en la medida en que, con un mismo modelo teórico, es viable llegar a un acuerdo en la interpretación de cuestiones morfológicas o sintácticas. Sin embargo, en las últimas décadas el PLN ha ido abordando cuestiones semánticas y más subjetivas donde el acuerdo interpretativo entre los anotadores cada vez es más complejo, incluso partiendo de los mismos modelos teóricos. En los últimos años, al tratar con temas tan subjetivos como la expresión de opiniones, emociones y sentimientos, o la detección de la ironía, se ha replanteado este modelo de calidad de un corpus anotado basado en el acuerdo entre anotadores.

En este contexto ha surgido un nuevo paradigma en anotación de corpora denominado “perspectivismo”.⁷ Este paradigma considera que el desacuerdo entre anotadores no es tanto un error como una fuente valiosa de información sobre el fenómeno anotado. El desacuerdo entre anotadores manifiesta formas diferentes de interpretar un mismo fenómeno lingüístico o (en nuestro caso) literario. Este desacuerdo es relevante porque, en primer lugar, está mostrando problemas en la anotación y por tanto temas relevantes para ser estudiados e investigados (¿a qué se debe el desacuerdo?, ¿es un fallo en el modelo teórico general, en las etiquetas, en el proceso?, ¿es un caso nuevo que obliga a replantear el modelo, la definición del fenómeno?, etc.); y en segundo lugar es relevante porque si entre las personas expertas no hay acuerdo, las herramientas computacionales derivadas del corpus deben aprender también que ese fenómeno se puede interpretar (es decir, anotar) de diversas maneras, que no hay una interpretación única (Cabitza *et al.* 2023).

El perspectivismo, por tanto, hace referencia a la posibilidad de anotar los mismos fenómenos y los mismos textos de diferentes maneras, desde supuestos, modelos o perspectivas diferentes. Toda anotación implica un proceso de interpretación, y se debe asumir que siempre puede haber diversas interpretaciones. Un corpus anotado debe ser consistente consigo mismo, es decir: una

7. Véase *The Perspectivist Data Manifesto* en <https://pdai.info/> (19 de septiembre de 2023)

misma persona debe anotar los mismos fenómenos de la misma manera a lo largo de todo el corpus. Pero no tiene por qué ser consistente con la anotación de otra persona.

Desde este punto de vista, una edición crítica enriquecida, en tanto que texto literario anotado, debe ser considerada siempre como una propuesta de análisis. La anotación fija una interpretación del texto que se presenta como referencia para otros estudios, pero no como una interpretación única y definitiva: se asume que puede haber (y debe haber) anotaciones y análisis alternativos que enriquezcan la interpretación. Si en otras áreas del PLN el perspectivismo es una opción, para la anotación del texto literario, dado su carácter plurisignificativo, es una necesidad. Piénsese, por poner un caso extremo (pero no extraño), las diferentes relaciones sintácticas y, con ello, las diferentes interpretaciones que se pueden establecer en estos famosos versos de Quevedo:

Cerrar podrá mis ojos la postrera
sombra que me llevare el blanco día,
y podrá desatar esta alma mía
hora a su afán ansioso lisonjera:

(Quevedo, *Amor constante más allá de la muerte*, vv. 1-4)⁸

Unidad

La unidad hace referencia a la unidad textual del corpus. Un texto literario anotado con información diversa a diferentes niveles de descripción lingüística puede llegar a ser un entramado complejo de ficheros e información. Pero no hay que perder nunca de vista la unidad del corpus: toda la anotación responde a un único texto y todos los elementos anotados deben estar relacionados.

Un poema, por ejemplo, anotado con información métrica y con información categorial debe estar creado de tal manera que, aunque sean aspectos relacionados con unidades diferentes (la sílaba para la información métrica y la palabra para la información categorial), quede claro que son rasgos de un mismo texto y que están relacionados.

Dada la diversidad de información y la heterogeneidad de unidades lingüísticas, establecer estas relaciones no es tarea fácil. En la siguiente sección se pondrán problemas derivados de este aspecto y vías de solución.

8. Jauralde Pou (1997) repasa los diferentes problemas interpretativos del poema y las lecturas clásicas. En concreto, para los versos 3-4, que considera “extremadamente difíciles” (pág. 94), destaca “la extraña utilización de *hora*; la concordancia de *lisonjera* y la referencia de *su*.” (pág. 95). Todos estos son problemas sintácticos. “Hora”, por ejemplo, suele considerarse sujeto de “podrá desatar” (Torre, 2004), pero no es una lectura evidente.

Sencillez y claridad

Precisamente por la complejidad que tiene un corpus anotado, se debe procurar que el modelo formal (el código, las marcas y las etiquetas utilizadas en la anotación) sea lo más claro posible para la persona que realiza la anotación manual. Con independencia del uso de editores de anotación como CATMA,⁹ INCEPTION¹⁰ u otros, un código claro permite a las personas modificarlo directamente. El objetivo es conseguir, en la medida de lo posible, un marcado claro, sencillo y explícito en la línea de las recomendaciones de código claro para lenguajes de programación como Python.¹¹

No quiero expresar con esto un rechazo absoluto al uso de editores de anotación. Más bien quiero expresar que el uso de editores de anotación debe ser siempre una opción del anotador, pero nunca una obligación. Los editores mediatizan la anotación y limitan la capacidad expresiva de los lenguajes de marcado.¹² Por ello siempre hay que mantener la opción de anotar directamente en código (con editores de código para asegurar la consistencia sintáctica y semántica), y ello implica que el marcado, el código, debe ser inteligible para un humano. Dado que en toda anotación se hará antes o después un trabajo manual, considero que hay que dar preferencia antes a la inteligibilidad del marcado que a su optimización computacional.

En resumen, de una edición digital enriquecida se espera que fije una interpretación de referencia del texto útil para los estudios computacionales, pero asumiendo siempre interpretaciones y anotaciones alternativas; que sea compatible con otros corpora y sistemas automáticos; que todos los niveles de anotación estén relacionados para mantener la unidad textual del corpus, y que el código sea, en la medida de lo posible, claro e inteligible para una persona experta.

Problemas formales en la anotación de un corpus multinivel.

Conseguir estas características implica resolver diversos problemas. En esta sección comentaré cuestiones derivadas de la formalización, la unidad y la claridad del código. A partir de ello, en la siguiente sección propondré un modelo general para poesía del Siglo de Oro.

9. <https://catma.de/> (19 de septiembre de 2023)

10. <https://inception-project.github.io/> (19 de septiembre de 2023)

11. Véase, por ejemplo, “Zen of Python” en *The Hitchhiker’s Guide to Python*: <https://docs.python-guide.org/writing/style/#zen-of-python>

12. El editor CATMA, por ejemplo, siguiendo las propias recomendaciones de TEI, separa el texto de la anotación en dos ficheros diferentes. El fichero con las etiquetas tiene referencias a las posiciones del fragmento de texto al que se refieren la etiquetas. Si bien este modelo es óptimo desde un punto de vista computacional, la anotación resultante resulta opaca para un humano.

Se comentó en la sección anterior que para conseguir la máxima interoperabilidad es necesario marcar el corpus con lenguajes de marcado de uso común y, si existe, seguir un estándar de anotación. En humanidades digitales, la opción más estandarizada es sin duda alguna el formalismo XML-TEI. Ahora bien, al afrontar la creación de un corpus multinivel, hay dos cuestiones generales a las que hay que dar respuesta: en primer lugar, cómo anotar aquellos fenómenos que no se pueden representar con XML-TEI o para los que hay un estándar alternativo; en segundo lugar, cómo resolver el problema de anidamiento que se puede producir entre los diferentes niveles de anotación.

Límites de XML-TEI

Si bien XML es un lenguaje de marcas muy flexible, que se puede adaptar a prácticamente cualquier tipo de información que se quiera representar, no siempre es la mejor opción. En anotación lingüística de corpora hay un caso en el que XML no es el lenguaje de marcas más utilizado: la sintaxis.

El modelo teórico más implantado hoy en sintaxis computacional es el modelo de dependencias. Mediante este modelo se marcan las relaciones de dependencia entre las palabras (lo que en la gramática tradicional se denominan funciones sintácticas: sujeto, objeto directo, etc.). La relación es siempre entre dos palabras, una es la palabra núcleo y otra la palabra dependiente. Todas las relaciones de dependencia entre las palabras de una oración forman al final el árbol sintáctico. Véase Fig. 1:

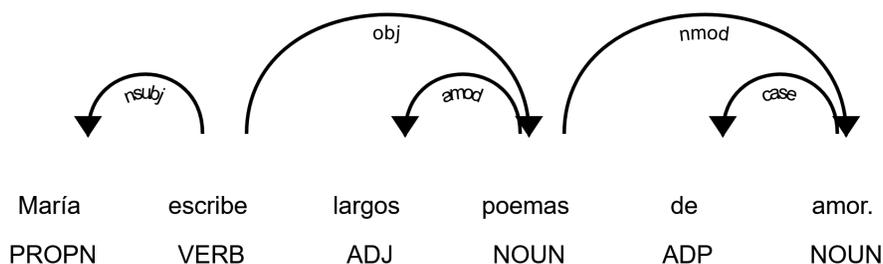


Fig. 1: Ejemplos de análisis sintáctico de dependencias.

Desde un punto de vista formal, un árbol de dependencias se representa como un grafo dirigido formado por un conjunto de vértices y un conjunto de arcos. Los vértices son las palabras de la oración y los arcos dirigidos son las relaciones de dependencia, que van siempre de un vértice (la palabra que actúa de núcleo) a otro vértice (la palabra dependiente de ese núcleo) (Jurafsky y Martin 2023).

El problema para la anotación de corpora con información sintáctica es que, si bien es posible representar grafos en XML, éste no suele ser el formalismo estándar. Hay otros modos más comunes de representar formalmente un grafo. En el caso concreto del análisis sintáctico, dentro del mundo del PLN el formato más común es el formato CONLL (Buchholz y Marsi 2006).¹³ La Tabla 1 muestra un ejemplo del análisis sintáctico del verso de Lope de Vega “Un soneto me manda hacer Violante” con un formato CONLL simplificado:

1	Un	Uno	DI0MS0	2	spec
2	soneto	soneto	NCMS000	5	obj
3	me	me	PP1CS00	4	obl
4	manda	mandar	VMIP3S0	0	sentence
5	hacer	hacer	VMN0000	4	obj
6	Violante	violante	NP00000	4	suj
7	.	.	Fp	4	f

Tabla 1: Análisis sintáctico de dependencias en formato CONLL (simplificado).

Como se puede ver, el formato CONLL es un formato tabular, donde las líneas representan cada palabra y las columnas su información morfosintáctica. La primera columna es el número identificativo de la palabra; y las siguientes tres columnas la información morfológica: el *token* tal y como aparece en el texto, su lema y la etiqueta categorial y morfológica. Finalmente, en las columnas 5 y 6 aparece la información sintáctica. La columna 5 es el número de la palabra de quien depende (su núcleo) y la 6 el tipo de dependencia.¹⁴

Así, la raíz de este árbol sintáctico es la palabra 4 (“mandar”), de la cual depende la palabra 6 (“Violante”) con una dependencia tipo “sujeto” (suj) y la palabra 5 (“hacer”) con una dependencia tipo “objeto” (obj), etc.

El siguiente código muestra este mismo grafo en anotación XML siguiendo las recomendaciones TEI:¹⁵

```
<graph type="directed" xml:id="RDG1">
<node n="1"><label>
  <w lemma="uno" pos="DI0MS0" msd="DI">Un</w></label>
</node>
<node n="2"><label>
```

13. Véase <https://universaldependencies.org/format.html> (19 de septiembre de 2023.)

14. En este caso, las etiquetas morfológicas y sintácticas están basadas en el modelo de Dependencias Universales. Véase <https://universaldependencies.org/> (19 de septiembre de 2023).

15. Véase el capítulo 19 de la guía TEI “Graphs, Networks, and Trees”: <https://tei-c.org/release/doc/tei-p5-doc/en/html/GD.html>

```

    <w lemma="soneto" pos="NCMS000" msd="NC">soneto</w></label>
</node>
<node n="3"><label>
    <w lemma="me" pos="PP1CS00" msd="PP">me</w></label>
</node>
<node n="4"><label>
    <w lemma="mandar" pos="VMIP3S0" msd="VMI">manda</w></label>
</node>
<node n="5"><label>
    <w lemma="hacer" pos="VMN0000" msd="VMN">hacer</w></label>
</node>
<node n="6"><label>
    <w lemma="Violante" pos="NP00000" msd="NP">Violante</w></label>
</node>

<arc from="#1" to="#2">
    <label>spec</label>
</arc>
<arc from="#2" to="#5">
    <label>obj</label>
</arc>
<arc from="#3" to="#4">
    <label>obl</label>
</arc>
<arc from="#4" to="#0">
    <label>sentence</label>
</arc>
<arc from="#5" to="#4">
    <label>obj</label>
</arc>
<arc from="#6" to="#4">
    <label>subj</label>
</arc>
</graph>

```

En este caso, primero se codifican los nodos del arco con la etiqueta <node> y luego los arcos dirigidos con la etiqueta <arc>, que especifica la relación entre dos nodos o palabras (@from y @to).

Como se puede observar, la codificación CONLL es más inteligible para un humano y es también óptima para su tratamiento computacional.

La cuestión, por tanto, a la hora de desarrollar una edición digital enriquecida multinivel, es decidir si para marcar la información sintáctica se utiliza XML, como presumiblemente se habrá utilizado en otros niveles, o utilizar otro formalismo diferente como CONLL, de uso común y considerado el estándar *de facto* en la comunidad científica.

Esto mismo podría ocurrir con otro tipo de información. Así, para la anotación multinivel de corpora lingüísticos hay propuestas en los dos sentidos. Por

ejemplo, PAULA-XML¹⁶ (Chiarcos *et al.* 2008) o SALT¹⁷ (Zipser y Romary 2010), al igual que TEI, proponen un único formalismo para marcar cualquier tipo de información en el corpus. El primer caso está basado en XML, mientras que el segundo está basado en grafos. De esta manera toda la información está unificada en un único formalismo. Sin embargo, otros corpora, como el corpus GUM¹⁸ (Zeldes 2017), utilizan un formalismo diferente para cada tipo de información o nivel de anotación, según sea el estándar o el formato más común. Así, en el corpus GUM la estructura general de cada documento está marcada en XML-TEI, pero luego utilizan CONLL para marcar la información sintáctica.

Dentro de los estudios literarios computacionales, esta opción de mantener el formato más usado para cada tipo de información o nivel de representación es la que se utiliza en BookNLP:¹⁹ un conjunto de herramientas de PLN entrenadas específicamente para procesar texto literario en inglés.²⁰ Por el contrario, algunos editores de marcado como CATMA²¹ o TEI:TOK,²² utilizan XML-TEI como formato nativo y cualquier tipo de información que se pueda anotar se representa con este formalismo único.

Con esta situación, no se puede considerar una opción mejor que otra. Si se opta por utilizar diferentes lenguajes de marcado, se corre el riesgo de perder la unidad del corpus y que éste quede como una acumulación de ficheros. Si se opta por utilizar XML para toda la anotación, puede resultar un código final oscuro; aparte del problema de anidamiento que se comentará después.

En cuanto a los límites de TEI, si bien cuenta con infinidad de etiquetas, no incluye toda la información que pueda interesar para los estudios literarios computacionales. Pero esto realmente no es un problema porque TEI ya está diseñado para que se pueda ampliar con la especificación de etiquetas propias, aunque no sea una práctica recomendada.²³

Anidamiento

Antes se comentó que la información que se puede anotar en un corpus multinivel es muy variada, desde información lingüística como lemas y categorías gramaticales.

16. <https://github.com/korpling/paula-xml> (19 de septiembre de 2023)

17. <https://corpus-tools.org/salt/> (19 de septiembre de 2023)

18. <https://gucorpling.org/gum/> (19 de septiembre de 2023)

19. <https://github.com/booknlp/booknlp> (19 de septiembre de 2023)

20. Junto al análisis lingüístico, realiza un análisis propio del texto literario como la detección automática de personajes y eventos narrativos, o la detección de estilo directo, entre otros aspectos.

21. <https://catma.de/> (19 de septiembre de 2023)

22. <http://teitok.corpuswiki.org/> (19 de septiembre de 2023)

23. Véase la sección “Customization” de la web de TEI: <https://tei-c.org/guidelines/customization/>. Para recursos en español, véase la web TTHUB <https://tthub.io/> (19 de septiembre de 2023) y en especial Allés-Torrent *et al.* (2022).

les, frases hechas, sintaxis, referencias a entidades (*named entities*), emociones y opiniones, relaciones anafóricas, estructuras argumentales y roles semánticos, significado léxico, etc. (Jurafsky y Martin 2023); hasta información literaria como los personajes de novelas y las relaciones entre ellos, intervenciones y estilo de discurso, eventos narrativos, métrica y ritmo, sentimientos, temas recurrentes, similitudes estilométricas, etc. (Piper *et al.* 2021, Schöch *et al.* eds. 2023, entre otros).

Si se opta por utilizar XML como único formalismo para representar información diversa en un único corpus, surge el problema del cruce de ramas, es decir, un anidamiento erróneo de los elementos XML.

La representación que hace XML de la información es una representación arbórea y anidada. De una etiqueta madre (la principal es *root*) pueden depender una, dos o más etiquetas, de las cuales a su vez pueden depender otras etiquetas, etc. Así, un texto marcado en XML es un árbol descendente, donde cada etiqueta de un nivel tiene anidadas las etiquetas del siguiente nivel.²⁴

Por ejemplo, para marcar la estructura básica de una novela en XML, TEI propone usar una etiqueta “text”, de la que depende una etiqueta “body”, de la que a su vez depende una o más etiquetas “chapter”, de la cual dependen una o más etiquetas “p” (párrafo), etc. Se forma, así, una estructura arbórea que se puede recorrer de manera descendente o ascendente.

Esta estructura arbórea establece relaciones de anidamiento entre etiquetas. Esto implica que si se introduce una etiqueta de apertura <A> y dentro otra etiqueta de apertura , no se puede cerrar la etiqueta <A> hasta que se haya cerrado la etiqueta anidada B. Si así se hiciera, el XML resultante estaría mal formado por un error de anidamiento y no se podría procesar.

Este error puede aparecer cuando se quiere anotar en un mismo fichero información a diferentes niveles de descripción, porque cada uno puede depender de unidades lingüísticas diferentes. Así ocurre, por ejemplo, si se quiere anotar en un poema información métrica e información categorial. La unidad básica de la métrica es la sílaba métrica, mientras que la unidad básica de las categorías gramaticales es la palabra (el lema). En principio, son unidades compatibles porque una palabra está formada por sílabas, por lo que se puede asumir que siempre que finalice una palabra, finalizará también una sílaba. El problema surge cuando aparece una sinalefa.

Efectivamente, con la sinalefa nos hallamos ante un problema de anidamiento: al unirse la última sílaba de una palabra con la primera sílaba de la siguiente palabra en una única sílaba métrica, no se puede representar ambas unidades con XML en un mismo fichero. Si se cierra la etiqueta de palabra, se debe cerrar la etiqueta de sílaba (que está anidada) y por tanto se pierde el límite de la sílaba métrica. Si no se cierra, no se puede cerrar la unidad palabra y quedarían dos palabras marcadas como si fueran solo una.

24. <http://www.w3.org/TR/xml/> (19 de septiembre de 2023)

En el verso de Lope de Vega mostrado en la Tabla 1 hay sinalefa entre las palabras “manda” y “hacer”. El siguiente ejemplo muestra las dos opciones de anotación (sílabas o palabras), pero no se pueden representar las dos a la vez:

```
<seg type="syllable">man</seg>
<seg type="syllable">daha</seg>
<seg type="syllable">cer</seg>

<w>manda</w>
<w>hacer</seg>
```

El estándar TEI propone algunas soluciones para poder tratar este problema (DeRose 2004).²⁵ En los siguientes párrafos las resumo brevemente con los siguientes versos de Lope de Vega como ejemplo. En este caso se produce un error de anidamiento entre la oración y el verso por el encabalgamiento del primer verso. Se muestra primero la anotación XML mal formada:

```
<l><s>Así Fabio lloraba. </s><s>Albania entonces </l>
<l>mirole, y quiso hablar, cerró los ojos, </l>
<l>y respondiolo lo demás la muerte. </s></l>
```

Como se ve, no se puede marcar el final del primer verso (<l>) sin antes marcar el final de la oración (<s>), pero no se puede marcar el final de la oración porque ésta finaliza en el tercer verso.

La primera solución es hacer la anotación en ficheros diferentes. Así, cada uno tendrá su XML y no habrá problema de anidamiento. Con esta opción, sin embargo, se pierde la unidad del corpus: habría que idear métodos para combinar la información de cada fichero de manera automática.

```
<l>Así Fabio lloraba. Albania entonces</l>
<l>mirole, y quiso hablar, cerró los ojos,</l>
<l>y respondiolo lo demás la muerte.</l>

<s>Así Fabio lloraba.</s>
<s>Albania entonces mirole, y quiso hablar, cerró
los ojos, y respondiolo lo demás la muerte.</s>
```

Una segunda opción es marcar las unidades no con etiquetas dobles (una de apertura y otra de cierre, tipo <A>), sino con etiquetas de elemento vacío <A/>, que aparecen solo una vez y marcan límites. La idea sería utilizar este tipo de etiqueta para marcar el inicio o el fin de la unidad que entra en conflicto con otra.

25. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/NH.html> (19 de septiembre de 2023)

En este ejemplo, en vez de poner una etiqueta `<s n=1>` para marcar el inicio de la oración y otra `</s>` para marcar su final, se podría marcar solo el inicio de cada oración en un texto con `<s n=1/>`. El final de la oración se puede inferir por la etiqueta de inicio de la siguiente oración, o por la marca de fin de texto. Al ser solo una etiqueta, no se produce conflicto en el anidamiento de etiquetas. Este método no es tan explícito como el uso de dos etiquetas de inicio y cierre, y en muchas ocasiones no se puede utilizar, pero es una solución elegante y útil para algunos casos.

```
<l><s n="1"/>Así Fabio lloraba. <s n="2"/> Albania entonces </l>
<l>mirole, y quiso hablar, cerró los ojos, </l>
<l>y respondiolo lo demás la muerte. </l>
```

Para la tercera opción es necesario que cada unidad tenga un identificador único (por ejemplo, si son oraciones, vale un simple `@n`, como `n=1` para la primera oración, `n=2` para la segunda, etc.). En este ejemplo, si una oración queda interrumpida por otra unidad, se cierra y se vuelve a abrir pero con el mismo identificador. De esta manera, gracias al identificador, queda marcado que la oración continúa:

```
<l><s n="1">Así Fabio lloraba.</s> <s n="2">Albania entonces </s></l>
<l><s n="2">mirole, y quiso hablar, cerró los ojos, </s></l>
<l><s n="2">y respondiolo lo demás la muerte. </s></l>
```

El problema de esta opción es que se duplican las etiquetas y queda una anotación algo redundante. Si el corpus es muy amplio, se pueden generar al final ficheros muy grandes.

La última opción es el llamado *stand-off Markup*. La idea es tener en ficheros separados el texto por un lado (sin ningún tipo de marca, el texto puro) y en otros la anotación, con tanto ficheros como tipos de anotación se quieran incluir. Para marcar a qué fragmento de texto corresponde cada etiqueta, se utilizan referencias entre ficheros como punteros, enlaces, etc. Esta es la opción que utilizan editores como CATMA. Computacionalmente la opción es óptima pero, como se ha comentado antes, el resultado es una anotación opaca para las personas pues, en vez del texto, aparecen números de referencia.

El siguiente ejemplo es el formato de CATMA. En cada etiqueta el texto está referenciado con dos números (punteros): el primero indica la posición en el fichero de texto del primer carácter del fragmento de texto y el segundo la posición del último carácter.

```
<content>Así Fabio lloraba. Albania entonces
mirole, y quiso hablar, cerró los ojos,
y respondiolo lo demás la muerte.</content>
```

```
<l><xi:include href="source.xml"
pointer="string-range(element(/1),0,35)"/></l>
<l><xi:include href="source.xml"
xpointer="string-range(element(/1),36,75)"/></l>
```

Como se puede observar, esta segunda opción es totalmente ininteligible para una persona. Eso sí, es muy eficiente desde un punto de vista computacional.

Qué solución adoptar depende de cada proyecto de anotación. En general, la opción 4 es la más eficiente desde un punto de vista computacional y la opción 2 la más elegante e inteligible, si bien no siempre es posible aplicarla. A partir de todos estos problemas, en la siguiente sección se mostrará un modelo general para la anotación multinivel de un corpus de poesía del Siglo de Oro.

Estudio de caso: un corpus multinivel de poesía del Siglo de Oro

En esta sección se aborda la anotación multinivel de un corpus de poesía del Siglo de Oro con el objetivo final de disponer de una anotación rica a modo de análisis de referencia que permita tanto profundizar en los estudios literarios como en el desarrollo de herramientas específicas de PLN para poesía española.

Los poemas están extraídos del *Corpus de sonetos del Siglo de Oro* desarrollado en el Proyecto ADSO (Navarro Colorado *et al.* 2016).²⁶ La edición de los textos es la modernización que realizó la Biblioteca Virtual Miguel de Cervantes,²⁷ más algunas correcciones y enmiendas. Además del texto, el corpus ya dispone de la anotación métrica realizada en el proyecto ADSO.

El modelo de anotación que aquí presentamos está pensado para poder marcar en principio cualquier tipo de información lingüística o literaria. Por ahora nuestro interés se centra en el estudio del hipérbaton como rasgo estilístico, para lo que se necesita marcar el corpus a tres niveles de descripción: métrico (el patrón métrico y rítmico de cada verso), léxico-categorial (la categoría gramatical de cada palabra) y sintáctico (las relaciones sintácticas entre las palabras, con independencia de su posición en el verso). En el futuro se espera poder anotar el corpus con otro tipo de información como entidades o lenguaje figurado.

El proceso de anotación, como es habitual, se desarrolla de manera semiautomática. Primero se realiza una anotación automática con herramientas de PLN, que luego es revisada, corregida y fijada de manera manual. En esta primera fase del proyecto estamos más interesados en fijar un análisis correcto que

26. <https://gplsi.dlsi.ua.es/proyectos/adso/> (17 de enero de 2024).

27. https://www.cervantesvirtual.com/portales/biblioteca_del_soneto/ (19 de septiembre de 2023).

en disponer de gran cantidad de texto anotado, por lo que por ahora se está trabajando con pocos sonetos.

Buscando la máxima interoperabilidad, el lenguaje de marcado utilizado es el XML siguiendo, en la medida de lo posible, las recomendaciones del estándar TEI. Esta elección no representa un problema para la anotación métrica y categorial, que está bien definida en TEI y para la que XML es un lenguaje de marcado apropiado.

Sin embargo, como se ha comentado en secciones anteriores, la representación de la información sintáctica en XML-TEI no es una representación estándar ni común. Para conseguir la máxima interoperabilidad es mejor utilizar el formalismo CONLL, al igual que hacen otros corpora multinivel como el corpus GUM (Zeldes 2017): este es el formato que utiliza la herramienta de análisis sintáctico utilizado y es más inteligible para la anotación manual. Para mantener la unidad del corpus, se utilizan identificadores, como se comentará luego.

Con la anotación métrica y léxico-categorial, ambas en XML, se puede producir un problema de anidamiento comentado anteriormente. Para solventarlo, seguiremos la opción de marcar cada nivel en ficheros diferentes (la opción primera de las cuatro indicadas en la sección anterior).

Para tratar este problema, en trabajos anteriores (Navarro Colorado 2019) se propuso marcar en un mismo fichero tanto la información métrica como la categorial, y así se materializó en el *Corpus general de poesía lírica castellana del Siglo de Oro*.²⁸ Con este objetivo, se desarrolló un modelo de anotación basado en XML-TEI. En este modelo, las palabras quedan marcadas con la etiqueta correspondiente (“<w>”), tanto al inicio como al final de cada una, que contiene además información sobre el lema de la palabra y su categoría gramatical. La separación silábica, sin embargo, no se marca con la etiqueta “<seg>”, que es la recomendada por TEI, sino con la barra vertical (“|”). Este formalismo de la barra vertical es recomendado por las guías directrices de TEI para marcar la separación silábica de palabras en diccionarios. De esta manera, la sinalefa queda explícitamente marcada con una etiqueta fin de palabra (“</w>”) que no está precedida por la barra vertical indicativa de fin de sílaba. Véase la sinalefa producida en “para andar” en este verso de Lope de Vega:

```
<l met="-.|-.|-.|5|-.|7|-.|" n="3">
  <w lemma="porque" type="CS">por|que|</w>
  <w lemma="para" type="SP">pa|ra|</w>
  <w lemma="andar" type="VMN0000">an|dar|</w>
  <w lemma="conmigo" type="PP1CSO0">con|mi|go|</w>
</l>
```

28. <https://github.com/bncolorado/CorpusGeneralPoesiaLiricaCastellanaDelSigloDeOro> (19 de septiembre de 2023)

Actualmente, sin embargo, no considero apropiado este modelo. Al incluir además la información métrica en la etiqueta de verso <l>, la anotación resultante es muy densa y resulta compleja para la anotación manual. Por otro lado, su procesamiento automático no resulta óptimo al mezclar dos tipos de etiquetas, unas basadas en XML (la etiqueta de verso y palabra) y otra no (la barra vertical para separar sílabas).

En la propuesta actual considero que es mejor tratar cada fenómeno en ficheros diferentes. Con ello se busca una anotación más modular, de tal manera que se puedan marcar los textos en paralelo a diferentes niveles, y que en cualquier momento se puedan incorporar nuevos niveles de anotación sin que afecte a los niveles ya marcados.

Así, siguiendo las recomendaciones TEI y aprovechando la experiencia en el desarrollo de otros corpora literarios como ELTeC (Burnard *et al.* 2022), se parte de un fichero central que contiene el texto puro (la edición crítica digital) marcada únicamente con información textual básica: los metadatos propios del encabezado TEI y la estructural (títulos y versos). El resto de niveles de representación se marca en ficheros diferentes.

El siguiente código es un ejemplo de la anotación métrica. Como se puede ver, es una anotación XML-TEI estándar: las sílabas se marcan con la etiqueta genérica “<seg>” más el valor “syllable”. El patrón métrico y rítmico con las etiquetas “<met>” y “<real>” respectivamente, más información sobre pausas y encabalgamientos. Para la justificación de este modelo de anotación, véase Navarro Colorado (2022):

```
<l source="#LopeVega.1063.1" met="--++-+++-"
    real="--++-+++-" enjamb="no">
<seg n="1" type="syllable" subtype="0"> Un</seg>
<w n="1"/>
<seg n="2" type="syllable" subtype="0"> so</seg>
<seg n="3" type="syllable" subtype="3"> ne</seg>
<seg n="4" type="syllable" subtype="0"> to</seg>
<w n="2"/> <pause type="1"/>
<seg n="5" type="syllable" subtype="0"> me</seg>
<w n="3"/>
<seg n="6" type="syllable" subtype="3"> man</seg>
<seg n="7" type="syllable" subtype="0"> da<w n="4"/> ha</seg>
<seg n="8" type="syllable" subtype="0"> cer</seg>
<w n="5"/>
<seg n="9" type="syllable" subtype="0"> Vio</seg>
<seg n="10" type="syllable" subtype="3"> lan</seg>
<seg n="11" type="syllable" subtype="0"> te</seg>
<w n="6"/>
<pause type="3"/>
</l>
```

El único dato ajeno a la métrica es el límite de las palabras. El final de cada palabra está marcado con la etiqueta <w>, pero como etiqueta simple <w/>. La razón por la que se incluye el límite de la palabra es para mantener la unidad del corpus: cada palabra (*token*) tiene un identificador único, que es utilizado en todos los niveles de anotación para poder relacionar unos ficheros con otros, independientemente del tipo de información anotada. Dado que estos identificadores actúan solo a modo de punteros, no es necesario hacer uso de la etiqueta de palabra “<w>” con apertura y cierre. Basta marcar el límite de cada palabra y su identificador con un elemento vacío “<w/>”, lo que evita el problema de anidamiento. Seguimos en este punto la segunda recomendación de TEI para evitar este problema comentado anteriormente.

Otro fichero diferente contiene la anotación categorial del poema. El siguiente fragmento de código muestra un ejemplo. En este caso también se siguen las recomendaciones TEI. Esta anotación es léxica, no hay marcas para oraciones ni sintagmas, por lo que no hay problema de anidamiento. Cada palabra está localizada con su identificador único, que es el mismo utilizado en el nivel métrico y sintáctico:

```
<l source="#LopeVega.1063.1">
  <w n="1" lemma="uno" pos="DI0MS0"> Un</w>
  <w n="2" lemma="soneto" pos="NCMS000"> soneto</w>
  <w n="3" lemma="me" pos="PP1CS00"> me</w>
  <w n="4" lemma="mandar" pos="VMIP3S0"> manda</w>
  <w n="5" lemma="hacer" pos="VMN0000"> hacer</w>
  <w n="6" lemma="Violante" pos="NP00000">Violante</w>
</l>
```

Finalmente, en otro fichero diferente se están haciendo pruebas de anotación sintáctica con el modelo teórico de las gramáticas de dependencias y el formalismo CONLL (véase un ejemplo en la Tabla 1 mostrada anteriormente). El identificador de cada palabra es el mismo utilizado en el resto del corpus. La información morfológica y categorial en este caso es redundante, porque ya aparece en el fichero del nivel categorial. Por ahora no se elimina para mantener el formato estándar. Lo relevante es la información sintáctica: las relaciones de dependencia entre las palabras que se han explicado en la sección anterior.

Con esto ya se podría llevar a cabo el marcado necesario para el estudio del hipérbaton como rasgo estilístico y su relación con el ritmo poético. Se dispone por un lado de la posición de cada palabra en el verso, de su patrón métrico y rítmico, de la categoría gramatical de cada palabra y demás información morfológica, y de las relaciones sintácticas de dependencias entre las palabras.

En estos momentos está definido el modelo y se están desarrollando las pruebas de anotación. Cada nivel de anotación tiene ahora sus propios problemas que serán tratados en su momento, sobre todo la anotación sintáctica que, sin duda, es la más compleja. Con todo, lo que aquí hemos mostrado es cómo

integrar toda esta información, más otra que se pudiera añadir al corpus, en un modelo modular pero unificado, claro en la medida de lo posible, y que asegure la máxima compatibilidad con otros corpora y herramientas de PLN.

Como crítica a este modelo podría argumentarse, y con razón, que es un modelo redundante, pues hay información repetida en diferentes ficheros (sobre todo el propio texto). Esto no es apropiado desde un punto de vista computacional porque hace que sea necesario mucho espacio en disco para almacenar el corpus. Precisamente el modelo de marcado *stand-off* se diseñó para optimizar el espacio y evitar redundancias.

Por ahora se ha preferido mantener la redundancia para facilitar la anotación del corpus en paralelo por diferentes personas y con información diferente. Repetir el texto en cada fichero permite al anotador centrarse en la información que está marcando, sin depender de otros ficheros. Así, por ejemplo, anotar la sintaxis se puede realizar con independencia del fichero con información categorial. Finalizada la anotación del corpus se podrá plantear un modelo *stand-off* para su almacenamiento final. Gracias a los identificadores de palabra y al fichero central con el texto puro se podrá automatizar este paso; pero esto queda ya como trabajo futuro.

Conclusiones

En este capítulo se ha mostrado la necesidad, tanto computacional como literaria, de desarrollar corpora literarios anotados a diferentes niveles de descripción lingüístico-literaria. Para considerar un corpus así anotado como una edición digital enriquecida que sirva de referencia para los estudios literarios, se ha argumentado que el modelo debe ser compatible con otros corpora y herramientas computacionales; que así mismo debería relacionar los diferentes niveles de análisis para mantener la unidad textual, y, finalmente, que sería deseable que fuera un modelo lo suficientemente claro como para permitir la edición manual del código. Por último, se han presentado algunos problemas formales y los procedimientos para solventarlos, mostrando así la viabilidad de desarrollar un corpus anotado de poesía del Siglo de Oro que cumpla con estas características.

Bibliografía

- ALLÉS-TORRENT, Susanna; Gabriel CALARCO, y Gimena DEL RIO RIANDE, “Edición y publicación de textos con TEI”, *TTHub. Text Technologies Hub: Recursos sobre tecnologías del texto y edición digital*, 2022 <<https://tthub.io/aprende/tutorial/edicion-y-publicacion-textos-tei/>>
- BUCHHOLZ, Sabine y Erwin MARSI, “CoNLL-X shared task on Multilingual Dependency Parsing”, *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-X)*, New York City, 2006, pp. 149-164.
- BURNARD, Lou; Borja NAVARRO COLORADO, Carolin ODEBRECHT y Martina SCHOLGER, “Collaborative creation of a multi-lingual literary corpus. Challenges and best practices for corpus design”, *COST Action Distant Reading Closing Conference*, Krakov/on-line, junio 2022.
- CABITZA, F., CAMPAGNER, A., BASILE, V., “Toward a Perspectivist Turn in Ground Truthing for Predictive Computing”, *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023 <<https://arxiv.org/pdf/2109.04270.pdf>>
- CHIARCOS, C., DIPPER, S., GÖTZE, M., LESER, U., LÜDELING, A., RITZ, J. & STEDE, M., “A Flexible Framework for Integrating Annotations from Different Tools and Tag Sets”, *Traitment automatique des langues* 49, (2008) pp. 271-293.
- DeRose, Steven, “Markup Overlap: A Review and a Horse”, *Proceedings of the Extreme Markup Languages*, Montréal, Canada, 2004.
- HAIDER, Thomas, “Metrical Tagging in the Wild: Building and Annotating Poetry Corpora with Rhythmic Features” *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 3715-3725, 2021 <<https://doi.org/10.18653/v1/2021.eacl-main.325>>
- HATZEL, Hans Ole; Haimo STIEMER; Chris BIEMANN y Evelyn GIUS, “Machine learning in computational literary studies” *it - Information Technology* (2023) <<https://doi.org/10.1515/itit-2023-0041>>
- IDE, Nancy y James PUSTEJOVSKY (eds.), *Handbook of Linguistic Annotation*, Dordrecht, Springer, 2017.
- JAURALDE POU, Pablo “Cerrar podrá mis ojos la postrera” *Revista de Filología Española*, LXXVII, 1/2 (1997).
- JURAFSKY, Dan y James H. MARTIN *Speech and Language Processing (3rd ed. draft)*, 2023 <<https://web.stanford.edu/~jurafsky/slp3/>>
- KESARWANI, Vaibhav; Diana INKPEN; Stan SZPAKOWICZ y Chris TANASESCU, “Metaphor Detection in a Poetry Corpus”, *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Vancouver, Canada, 2017, pp. 1-9.
- McSHANE, Marjorie y Sergei NIRENBURG, *Linguistics for the Age of AI*, Cambridge, Massachusetts, MIT Press, 2021.
- MEISTER, Jan Christoph, “Poetry, Phenomenon and Phenomenology” en BORRIES, Anne-Sophie; Petr PLECHÁČ y Pablo RUIZ FABO (eds.), *Computational Stylistics in Poetry, Prose, and Drama*, Berlin/Boston, de Gruyter, 2022.

- Navarro Colorado, Borja; Ribes-Lafoz, María and Sánchez, Noelia “Metrical Annotation of a Large Corpus of Spanish Sonnets: Representation, Scansion and Evaluation”, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* Portorož, Eslovenia, 2016.
- NAVARRO COLORADO, Borja, “A Metrical Scansion System for Fixed-Metre Spanish Poetry”, *Digital Scholarship in the Humanities* 33, 1 (2018a), pp. 112-127
- NAVARRO COLORADO, Borja, “On Poetic Topic Modeling: extracting themes and motifs from a corpus of Spanish poetry”, *Frontiers in Digital Humanities. Computational Linguistics and Literature*, (2018b), DOI 10.3389/fdigh.2018.00015
- NAVARRO COLORADO, Borja “Por un análisis distante y profundo: un corpus piloto de la poesía lírica castellana del Siglo de Oro” *Revista de poética medieval*, 33 (2019).
- NAVARRO COLORADO, Borja, “Beyond the metre: formalization of linguistic-rhythmic features for a computational analysis of the Spanish poetry”, *Calderón, R and Python: New Methods and Digital Tools for Quantitative Analysis of Theatre, Verse Drama and Poetry*, Vienna, University of Vienna, 2-3 junio 2022.
- PIPER, Andrew; Richard J. So y David BAMMAN, “Narrative Theory for Computational Narrative Understanding”, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 298-311. <<https://doi.org/10.18653/v1/2021.emnlp-main.26>>
- PUSTEJOVSKY, James y Amber STUBBS, *Natural Language Annotation for Machine Learning*. O’Reilly Media, Inc., 2012.
- DE LA ROSA, Javier; Álvaro PÉREZ; Laura HERNÁNDEZ; Salvador ROS y Elena GONZÁLEZ-BLANCO, “Rantanplan, Fast and Accurate Syllabification and Scansion of Spanish Poetry”, *Procesamiento del lenguaje natural* 65 (2020), pp. 83-90
- SCHÖCH, Christof; Julia DUDAR, Evgeniia FILEVA, eds., *Survey of Methods in Computational Literary Studies* Version 1.1.0, Trier, CLS INFRA, 2023 <<https://methods.clsinfra.io>>, DOI: 10.5281/zenodo.7892112.
- THE TEI CONSORTIUM, *TEI P5: Guidelines for Electronic Text Encoding and Interchange 2023* <<https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>>
- TORRE, Esteban, “«Cerrar podrá mis ojos...»: Paráfrasis, métrica y hermenéutica” *Rhythmica*, 2, (2004), pp. 235 - 250 <<https://doi.org/10.5944/rhythmica.6409>>
- Zeldes, Amir, “The GUM Corpus: Creating Multilayer Resources in the Classroom”, *Language Resources and Evaluation* 51, 3 (2017), pp. 581-612.
- Zipser, F. y L. Romary “A model oriented approach to the mapping of annotation formats using standards”, *Proceedings of the Workshop on Language Resource and Language Technology Standards, LREC 2010*, Malta, 2010 <<http://hal.archives-ouvertes.fr/inria-00527799/en/>>