

# 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.