

ROCHEBOUET, Anne (ed.). Le texte médiéval à l'épreuve du numérique.
Médiévales, 73, 2017. ISBN 978-2-84292-801-8

<https://journals.openedition.org/medievaux/8157>

doi: 10.4000/medievaux.8163

Reseña realizada por
Marta LÓPEZ IZQUIERDO
Université Paris 8
martali.2011@gmail.com

Editado por la medievalista francesa Anne Rochebouet, el volumen se compone de 9 artículos: una introducción de la editora y 8 contribuciones que emanan de los trabajos presentados en la Jornada de estudios celebrada en junio de 2016 en la Universidad de Versailles Saint-Quentin.

Como explica la editora en su introducción, la jornada se propuso dar a conocer herramientas digitales creadas en diversos proyectos de Humanidades para favorecer su reutilización, así como la de los datos obtenidos, en nuevos proyectos de otros investigadores y paliar de esta manera la escasez de medios técnicos, humanos o financieros en este ámbito de la investigación. El conjunto se centra en el estudio de datos textuales, tanto sobre su contenido semántico como sobre el soporte en que aparecen. Sin embargo, la visualización de los textos en el medio digital adolece de una verdadera reflexión epistemológica sobre los diferentes estatutos del texto de partida: transcripción, facsímil, edición crítica, testimonio, obra, y tiende a una homogeneización engañosa tras una misma apariencia en la pantalla del ordenador.

Rochebouet denuncia con razón la falsa impresión de neutralidad o de objetividad frente a la mediación de las pantallas y los ordenadores, puesto que toda migración de soporte implica una operación de transposición y supone una serie de elecciones y de interpretaciones.

La utilización masiva de las digitalizaciones de documentos, disponibles en la actualidad en la mayoría de los repositorios de las principales bibliotecas, ha hecho emerger una *cultura del facsímil* así como nuevas prácticas de trabajo que, a juicio de la editora, no se han visto acompañadas de una reflexión sobre la naturaleza de las intervenciones subyacentes.

Frédéric Duval dedica su contribución, titulada “Pour des éditions numériques critiques. L'exemple des textes français” (pp. 13-29), a las ediciones críticas digitales de textos franceses medievales. Una primera constatación da cuenta del escaso número de ediciones digitales *nativas* en el ámbito del medievalismo francés. A su vez, el autor alerta contra el peligro de seguir ciegamente prácticas digitales usuales sobre todo en el mundo anglosajón, que no siempre están en adecuación con la naturaleza de la realidad textual medieval ni toman en cuenta la tradición ecdótica de la romanística, al identificar la edición digital exclusivamente con un modelo de edición documental, más fácil de realizar y de interrogar con motores de búsqueda automatizados. Ahora bien, la edición documental puede servir de base o etapa preliminar para la realización de una edición crítica, por lo que ambos modelos no deben entenderse como necesariamente excluyentes. La edición digital debe desarrollar su principal ventaja: la modularidad y la posibilidad de proponer diversos recorridos de consulta y lectura, sin renunciar al trabajo de acercamiento al texto del autor. Asimismo, la edición digital documental de los testimonios disponibles facilitaría enormemente la labor crítica permitiendo una *collatio* automatizada. Dentro de la dimensión histórica que es propia de Filología, la edición digital científica promovida por Duval puede combinar, de manera enriquecedora, la perspectiva documental y autorial.

Las dificultades con que se enfrenta la edición crítica digital para existir en su plenitud proceden, según el autor, de las carencias técnicas en el proceso de codificación: el lenguaje XML-TEI hoy comunmente aplicado en la labor de edición digital no permite la construcción de aparatos críticos dinámicos, superponibles, que ofrezcan visibilidad a la vez del documento (o de los documentos) y del texto. Las innovaciones técnicas, advierte Duval, no podrán ver la luz sin una actividad editora que promueva y proponga soluciones para asegurar el futuro de la edición crítica digital y garantizar así la adecuada transmisión de los textos antiguos y medievales en toda su complejidad.

Graziella Pastore presenta en su artículo la base prosopográfica Omeka y la etapa inicial de su construcción [“L'édition numérique du *Livre de justice et de plet* (Élec), point de départ de la constitution d'une base prosopographique (Omeka)”, pp. 31-45]. El proyecto parte de la necesidad de mejorar la intercomunicación de los archivos prosopográficos, cada vez más numerosos, y su manejo por parte de los numerosos usuarios y productores de datos

(investigadores, documentalistas, instituciones). Su objetivo inicial ha consistido en proponer un sistema de publicación y de gestión de ficheros de autoridad al alcance de investigadores con un nivel bajo de competencias informáticas para la elaboración de una base de datos prosopográficos. Por su gran adaptabilidad, este sistema es exportable a otros campos de estudio histórico.

Se escogió el formato de metadatos EAC-CPF (Encoded Archival Context-Corporate Bodies, Persons and Families), en lenguaje XML, por su gran difusión y por la complejidad estructural que permite en la descripción de personajes. Para la visualización, el sistema de gestión de contenidos (CMS) Omeka presenta grandes facilidades en la creación de un sitio web y la gestión de varios perfiles de usuarios. Para la fuente de los datos, se optó por un corpus reducido de personajes (unos quince entre merinos, oficiales y consejeros del reinado de Luis XI) citados en el *Livre de justice et de plet*, compilación jurídica de la segunda mitad del siglo XIII, disponible en edición digital en la colección Élec¹.

El sitio que aloja el resultado de este proyecto se encuentra en Huma-Num (Consortio Humanités Numériques)², y está enlazado con la edición en línea del texto medieval. Los contenidos del sitio están organizados en dos colecciones: 1) persona (nombres de personas citadas), 2) bibliografía (referencias que contienen información sobre las personas citadas).

El proyecto descrito ha permitido, en un plazo razonablemente breve, gestionar la cadena editorial de noticias de autoridad, desde su creación hasta su visualización, difusión en línea, y descarga de ficheros. La autora apunta algunas modificaciones susceptibles de mejorar el manejo de la base, así como sus posibilidades de visualización o la ampliación del corpus de partida.

El artículo de Sergio Torres Aguilar [“La reconnaissance des entités nommées dans les bases numériques de chartes médiévales en latin: le cas du *Corpus Burgundiae Medii Aevi* (Xe – XIIIe siècle)”, pp. 47-65] se interesa por los métodos que permiten explotar a gran escala la ingente masa de manuscritos medievales digitalizados que está disponible en la actualidad. Dentro de esta perspectiva, Torres Aguilar ofrece una propuesta de estructuración automatizada del corpus CBMA (con más de 29.000 actas y 300 cartularios), basada en el reconocimiento de las *entidades nombradas* (inglés, *named entity*; francés, *entité nommée*), por el cual se puede atribuir un nombre específico y una clasificación a todos los elementos referenciales presentes en el texto, relativos a una persona, un lugar o una organización. La identificación correcta de las entidades con unidades básicas de información (quién, dónde, cuándo) permite una interrogación dinámica de la base de datos a gran escala. Gracias a la estructuración de los textos, se puede obtener un gran volumen de información sobre la historia

¹ Véase Colección Élec, <http://elec.enc.sorbonne.fr/justiceetplet/>.

² Véase Huma-Num (Consortio Humanités Numériques) <http://justiceetplet.huma-num.fr>.

de un individuo o de una comunidad, generar árboles de relaciones sociales o de jerarquías especiales, o bien conocer la evolución de nombres, usos, conceptos.

Desde el punto de vista formal, el reconocimiento de nombres de entidades se funda en indicios sintácticos y morfológicos, a la vez que contextuales, que permiten distinguir en el texto las entidades que pertenecen a la clase y le atribuyen una etiqueta. Se analizan a continuación los porcentajes de acierto obtenidos según los tipos de entidades (simples o compuestas, de persona/geográficas/organizaciones). La aplicación del modelo sobre corpus procedentes de otras regiones (Castilla, Lombardía, Isla de Francia y Sur de Inglaterra) ofrece unos porcentajes de acierto ampliamente satisfactorios. Sin embargo, el porcentaje de fallos aumenta en los siglos periféricos del corpus (X y XIII) con respecto a los siglos centrales (XI y XII).

La interpretación de estos resultados no puede hacerse, indica Torres Aguilar, sin conocer los estudios de antroponimia medieval que coinciden en reconocer una evolución similar en toda Europa a partir de finales del siglo X. Así, mientras la presencia del nombre doble, bastante estable durante los siglos XII y XIII ofrecía resultados de alta fiabilidad, la situación del siglo X, en que domina el nombre simple, o del siglo XIII, en que las formas se hacen cada vez más complejas, se corresponde con una eficacia menor del modelo.

La propuesta de Torres Salvador combina por consiguiente la introducción de técnicas y metodologías digitales con la observación de las especificidades históricas del documento, dentro de un trabajo necesariamente interdisciplinar.

Dominique Stutzmann, Jean-François Moufflet y Sébastien Hamel dedican el artículo siguiente “La recherche en plein texte dans les sources manuscrites médiévales: enjeux et perspectives du projet HIMANIS pour l’édition électronique”, pp. 67-96] al proyecto HIMANIS (*Historical Manuscript Indexing for user-controlled Search*), que permite indexar el texto de los registros de la Cancillería real francesa (siglos XIV-XV) a partir de la imagen obtenida por escaneo. La masa de información disponible en los registros, sobre todo a partir del siglo XV, es tal que no ha podido ser estudiada más que muy parcialmente hasta la fecha. El proyecto HIMANIS se propone facilitar su estudio a través de la indexación automática del texto, para lo cual se ha realizado un marcado de los metadatos con TEI y se ha aplicado un programa de reconocimiento y transcripción automática de escritura manuscrita (HTR, *Handwritten Text Recognition*) de una parte de los registros, para alcanzar una doble meta: tecnológica (búsqueda por palabras clave, indexación, identificación de copistas y búsqueda por negociación de contenido) y en humanidades (investigación histórica y lingüística dentro de los documentos, análisis de datos masivos, análisis de redes, lingüística de corpus). Varios aspectos han sido desarrollados con estos objetivos: la incorporación de referenciales lingüísticos (diccionarios de frecuencias a partir de ediciones existentes), el reconocimiento de entidades nombradas (lugar, persona, conceptos), y finalmente, la granularidad de acceso, que exige la identificación de cada zona de imagen con cada una de las actas repertoriadas, tarea que

escapa a la automatización. La indexación así obtenida permite situar el corpus dentro del universo de datos abiertos y enlazados, con posibilidades de visualización y de explotación diversas. Los autores concluyen su artículo con dos *caveat*, generalizables al conjunto de las Humanidades Digitales: el riesgo de repetir evidencias o el de llegar a conclusiones falsas.

Xavier-Laurent Salvador [“Isilex, un serveur et gestionnaire de contenus collaboratifs REST XQuery au service des humanités numériques: l’exemple de la Bible historique”, pp. 97-114] es autor de una herramienta de codificación y puesta en línea de corpus automatizados, Isilex, especialmente pensada para trabajar con corpus bíblicos medievales (la *Bible historique* francesa), caracterizados por su polifonía (traducciones en varias lenguas, inserción de citas de autoridades, etc.). Su propuesta se inscribe en una perspectiva de estudio del dato en su contexto o *small data*. La arquitectura REST (*Representation State Transfer*) permite asociar una URL directamente a una reserva de representaciones de datos que se forman gracias a la propia URL sin necesidad de pasar por la consulta de una colección de datos. Así, la obra no es el producto de la labor de un autor o de un editor, sino el resultado de las operaciones de selección del lector que construye un texto a la medida de sus intereses. Tanto XML, utilizado para la estructuración de los datos, como XPath, para localizar datos dentro de una arborescencia XML, se presentan como bloques estáticos que no pueden dar respuesta a las búsquedas dinámicas de un lector del texto medieval. Los sistemas gestores de bases de datos asociados al lenguaje de búsqueda XQuery permiten superar esta dificultad. Isilex es una herramienta polifuncional: servidor web basado en BaseX, un CMS (*Content Management System*) de contenidos colaborativos XML/XQuery, que permite la construcción de enciclopedias, la puesta en línea de corpus y el cruce automático de datos gracias a un servidor rest y al conjunto de funciones full-text de XQuery que incluye. Por otro lado, acepta nuevas aplicaciones y extensiones, por lo que se presenta, concluye Salvador, como una solución ampliamente adaptable para la investigación innovadora en las ciencias del texto, que, más allá de la arquitectura del documento, integran nuevas cuestiones relativas a la información, la comunicación y el dato.

En “Les outils CapiTainS, l’édition numérique et l’exploitation des textes” (pp. 115-131), Thibault Clérice présente el proyecto, del que es co-autor, junto a Bridget Almas, CapiTainS, conjunto de herramientas *open source*, para la edición de textos en libre acceso que cumpla con los requisitos de científicidad, normalización y almacenamiento. En su origen, el proyecto se concibió para producir textos compatibles con el modelo CTS (*Canonical Text Services*), que reposa sobre un sistema de identificadores con cinco elementos: un dominio, un identificador *textgroup*, un identificador *work*, un identificador *version* y un identificador *passage*. Todos ellos, menos el último, responden a una jerarquización de los metadatos de un texto. Se suma a este nivel un sistema de cita en el interior del texto por medio de etiquetas TEI, XML y XPath.

La interfaz de acceso se construye en CapiTainS gracias a un conjunto de herramientas *open source* y libres siguiendo normas reconocidas y compartidas, pero dejando la posibilidad de personalizar la interfaz de usuario. La interfaz máquina (API, *Application Programming Interface*) lleva a cabo la comunicación entre máquinas: la aplicación *Nautilus*, desarrollado dentro del proyecto, permite utilizar el protocolo CTS y DTS (*Distributed Text Services*) y prevé incluir una interfaz Sparql (para la interrogación de metadatos RDF, *Resource Description Framework*) y una interfaz OAI-PMH (para la interrogación de catálogos). La interfaz de usuario es personalizable vía la aplicación Nemo, del mismo paquete CapiTainS.

La meta principal del proyecto es sin embargo la puesta a disposición de recursos textuales para su reutilización sin que se pierda su formato estándar. Para preservar su calidad académica, los recursos deben ser citables (fuente, momento de producción, producción del recurso digital), como en el caso de los DOI (*Digital Object Identifiers*) y conformes a una serie de normas preestablecidas, generalmente TEI-XML. CapiTainS ha desarrollado los sistemas Hook y HookTest para dar respuesta a estas dos necesidades.

En definitiva, el conjunto de herramientas del proyecto CapiTainS pone a disposición de los investigadores la creación de plataformas de lectura, la estructuración y visualización de datos o metadatos con fines y formas diferentes, la concepción de la producción y la reexplotación de datos, y pretende así simplificar y abaratar el trabajo de los filólogos y los historiadores.

Anne Salamon dedica su contribución a presentar el portal digital HU 15 (a partir de un corpus de historias universales inéditas en francés del siglo XV) y de manera más general, a reflexionar sobre el papel del filólogo en la sociedad y en el mundo universitario de hoy: ¿cómo puede renovar su labor con/frente a los textos a través de las herramientas digitales ya existentes? La reflexión de Salamon pasa en primer lugar por el concepto de *remediación*, tomado de Bolter y Grusin: cada nuevo medio incorpora una parte de las características del medio que lo precede. Frente a la *inmediación* (el medio se borra hasta hacerse transparente para el espectador/receptor), se sitúa la *hipermediación*, cuyo objetivo es dar a ver el medio. El nacimiento de la edición digital se acompaña de un movimiento de desconfianza hacia el nuevo medio: responde a un efecto de hipermediación, en que el medio utilizado prima sobre la calidad del texto editado. Así, las ediciones impresas siguen gozando de mayor prestigio que las electrónicas. Paradójicamente, lo que impulsa las ediciones digitales es un deseo de mayor transparencia, una mayor cercanía al texto *real*, sin mediación, aunque en realidad el nuevo medio no ha supuesto un nuevo marco teórico de comprensión y concepción de lo que son los textos. La ausencia de modelo de edición electrónica es una prueba, según Salamon, de que el nuevo medio no se ha liberado aún de los medios antiguos. Ahora bien, el nuevo medio renueva nuestra manera de leer el contenido.

Así pues, se buscó un tipo de edición digital que se alejara del formato comunmente utilizado y que no necesitara innovaciones tecnológicas propias. Salamon parte de la corriente de la *nueva codicología*, en que el códice se convierte en el objeto central de estudio. El proyecto HU15 pretende relacionar el texto con su contexto, proporcionando un acceso virtual a los elementos materiales, visuales y textuales característicos de los manuscritos. El sitio web busca *remediar* el aspecto material del manuscrito, no a través de una aproximación mimética (fac-símil) sino hipermediática. Permite al lector, gracias a los índices variados de que dispone la edición, realizar una lectura temática de las distintas versiones según sus intereses. Salamon concluye evaluando los logros del proyecto en términos de terreno de experimentación y proponiendo una prolongación de su contenido, así como de su metodología.

El último artículo es obra de Davide Gherdevich [“PolimaWiki: un site contributif pour l'étude du pouvoir des listes au Moyen Âge”, pp. 149-168], en el que presenta el proyecto de sitio web colaborativo PolimaWiki, dedicado al poder de las listas en la Edad Media. El proyecto se propone contribuir al conocimiento de la cultura medieval de lo escrito a través del estudio de las listas, definidas en términos sintácticos como un texto construido con elementos paratácticos (enumeraciones, listas, catálogos, inventarios, etc.). Los investigadores implicados en el proyecto buscaban un espacio de trabajo colaborativo para alojar documentos, poder intercambiar información y elaborar conjuntamente hipótesis de trabajo. La creación de la plataforma digital POLIMA pretende dar respuesta a estas demandas por medio de la adopción de una tecnología Open Source de tipo wiki. El programa seleccionado ha sido MediaWiki, CMS de la enciclopedia Wikipedia. Además de responder a los requisitos de almacenaje y de espacio colaborativo, MediaWiki dispone de varios módulos de extensión, entre los que se ha utilizado Semantic MediaWiki, que permite añadir etiquetas semánticas a las páginas de una wiki y con ellas realizar búsquedas semánticas y compartir contenidos como una base de datos. Para la creación de las páginas de la wiki, se ha utilizado la extensión Page Form, que permite añadir y editar páginas. Gracias a estos dispositivos, el utilizador puede abrir un nuevo “formulario” estructurado en cuatro partes: identificar, describir, depositar y documentar. Todos los documentos reunidos constituyen una base de datos consultable en línea, que puede recibir ampliaciones o correcciones posteriormente, vía un moderador. Las intervenciones de los investigadores exteriores al proyecto son también posibles. Por otro lado, es siempre posible exportar los datos y resultados de la búsqueda en formato RDF y XML. Finalmente, los manuscritos y documentos pueden ser localizados visualmente sobre un mapa (Google Earth) gracias a la extensión Semantic Maps.

Las contribuciones editadas en este número de la revista *Médiévales* muestran los diversos desafíos con que deben enfrentarse los estudios medievales ante el *digital turn*. En el volumen, alternan de manera eficaz enfoques más técnicos con problemáticas propias de las humanidades, mostrando los diversos niveles de reflexión e investigación que deben asociarse

en el campo de las Humanidades Digitales. Los interrogantes y las soluciones evocadas en el conjunto del volumen ayudan a situar la actualidad de la investigación en torno a la edición digital científica en Francia y enriquecen de manera muy necesaria un panorama ampliamente dominado por las producciones anglófonas.

El número de la revista se concluye con tres contribuciones dedicadas a la Literatura Medieval Europea y cuatro reseñas de monografías dedicadas a la Historia Medieval Francesa, fuera ya de la temática de las Humanidades Digitales, por lo que no las reseñamos aquí.