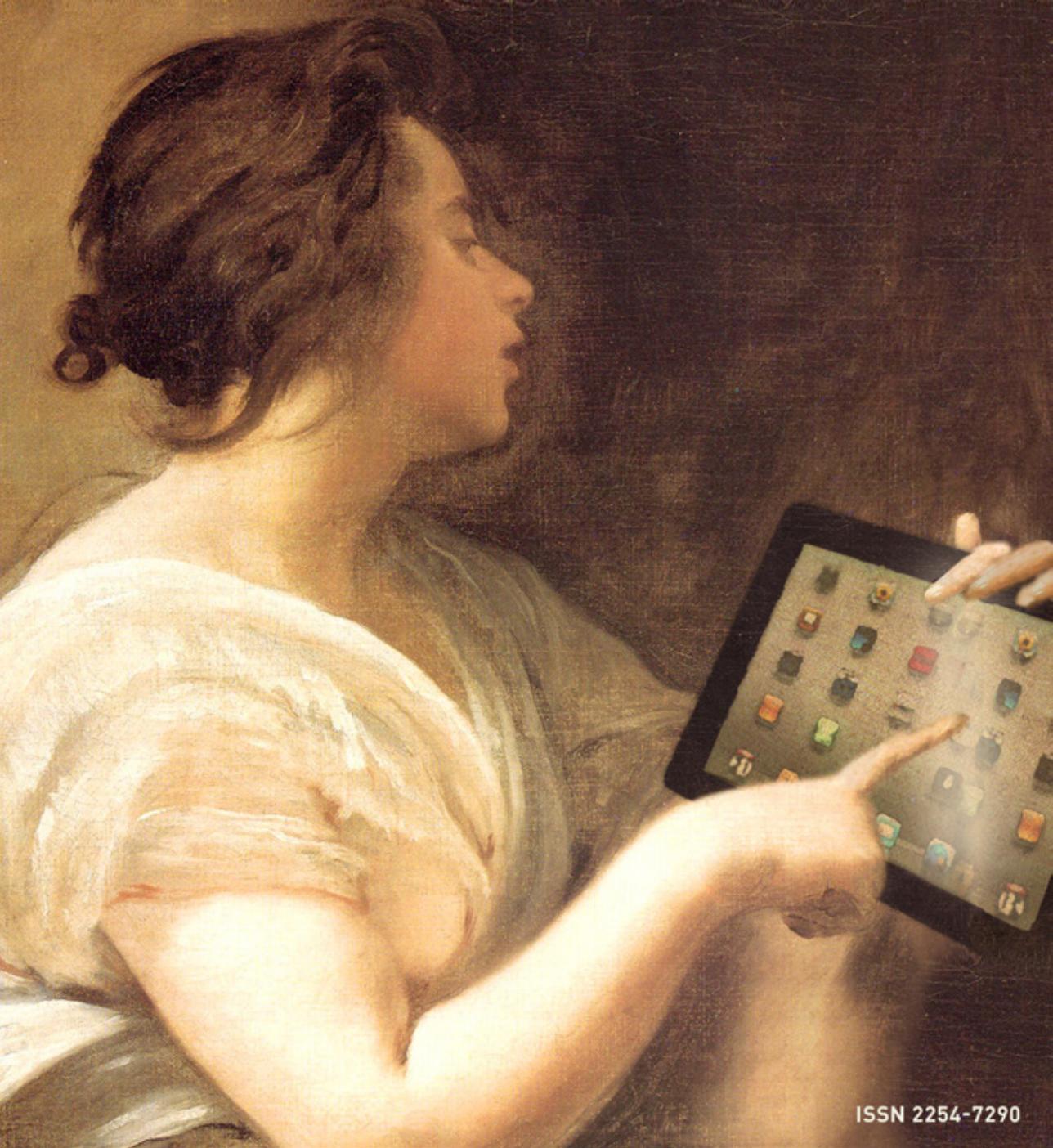




HUMANIDADES DIGITALES: DESAFÍOS, LOGROS Y PERSPECTIVAS DE FUTURO

SAGRARIO LÓPEZ POZA Y NIEVES PENA SUEIRO (EDITORAS)



Humanidades digitales: desafíos, logros y perspectivas de futuro



Sagrario López Poza
y Nieves Pena Sueiro
(editoras)

Janus. Anexo 1 (2014)

ISSN 2254-7290

Diseño de la cubierta: Paula Lupiáñez López

Edita:

SIELAE (Seminario Interdisciplinar para el Estudio de la Literatura Áurea Española).
Facultad de Filología. Universidade da Coruña. A Coruña (España)

© 2014 Los autores (cada uno de su trabajo)

© De esta edición: SIELAE-JANUS

Reservados todos los derechos.

SIELAE

<http://www.bidiso.es/sielae>

Maquetación: Dosgraphic, s.l.

ÍNDICE

Sagrario López Poza Presentación	7
José Luis Canet Reflexiones sobre las humanidades digitales	11
John Griffiths Musicología, informática y la vihuela en el siglo XXI	21
Paul Spence Centros y fronteras: el panorama internacional de las humanidades digitales..	37
	
Diego Agudelo Grajales Proyectos sociales y de aula a partir del aprendizaje colaborativo virtual. Experiencias desde la tutoría en Diplomado de AUSJAL.....	63
Montserrat Amores El buscador de GICES XIX, herramienta digital sobre el cuento español del siglo XIX	79
Gonzalo A. Aranda Corral, Joaquín Borrego Díaz y Juan Galán Páez Extracción y organización del conocimiento de etiquetados. Aplicación a etiquetados en repositorios digitales sobre arte	87
Consolación Baranda y Eduardo Rodríguez Red ARACNE: retos y objetivos de un proyecto de coordinación en letras hispánicas digitales.....	101
Gidor Bilbao y Ricardo Gómez Textos antiguos vascos en Internet.....	111
Jean-Marc Buiguès Indexación y códigos en bases de datos bibliográficas.....	123

Paloma Centenera Centenera y María Luisa Díez Platas

El proceso de desarrollo de una biblioteca digital y su apertura e integración en el recolector *Hispana*: el caso de la *Biblioteca Digital Ovidiana Abierta*..... 137

Daniel Escandell Montiel

Tecnoestética de la blogoficción: imitación y suplantación del bloguero en la narrativa del avatar 151

Carme Fernández Pérez-Sanjulián, M^a Antonia Pérez Rodríguez, Eduardo Rodríguez López y Ánxeles Saavedra Places

Recursos para a clasificación da produción editorial na Galiza durante a etapa franquista: diseño e alimentación da base de datos 161

Carlota Fernández Travieso

Base de datos sobre Inventarios y Bibliotecas del Siglo de Oro (IBSO). Utilidad y posibilidades 175

Carolina Ferrer

El canon literario hispanoamericano en la era digital 185

Juana María González García

El proyecto *Epístola*: edición digital de los epistolarios de la Edad de Plata 197

Elena González-Blanco García, Clara Isabel Martínez Cantón, María Dolores Martos Pérez y María Gimena del Río Riande

Una propuesta de integración del sistema de formularios de bases de datos MYSQL con etiquetado TEI: ReMetCa, Repertorio digital de la métrica medieval castellana..... 209

M^a Isabel Hidalgo Urbaneja

Análisis crítico de los discursos y narrativas digitales. *The Gallery of Lost Art* –Tate– como estudio de caso 221

M^a Eugenia Iglesias Moreno, Pilar Azcárate Aguilar-Amat y Sonia Sánchez Cuadrado

Primera aproximación para la extracción automática de *Entidades Nombradas* en corpus de documentos medievales castellanos 229

Xurxo Insua, Mireia Freixa y Teresa-M. Sala

Experiencias de e-research y divulgación digital en historia del arte: GracmonDocs y Viquimodernisme 239

Raquel Jimeno Revilla

Los clubes del libro en la era digital: nuevas estrategias y aspectos sociales. El caso de Círculo de Lectores (1998-2012)..... 247

Sonia Lavandeira Rojo

Publicidad digital 257

Laura Llanos Casado y Milka Villayandre Llamazares

Nuevas tecnologías y escritura académica 263

Álvaro Llosa SanzDe cómo Ramón Gómez de la Serna se convirtió en un fan de Twitter:
un ejemplo de textualidad digital..... 277**Sagrario López Poza y Ángeles Saavedra Places**Recursos digitales ofrecidos por el SIELAE para el estudio del Siglo de Oro.
Creación, gestión y evolución de BIDISO..... 285**Ana Mancera Rueda y Ana Pano Alamán**Las redes sociales como corpus de estudio para el Análisis del discurso
mediado por ordenador 305**Patricia Meilán Jácome y Fátima Díez Platas**La *Biblioteca Digital Ovidiana* y el patrimonio bibliográfico español 317**Isa Bernardino Mestre**Literatura digital o la reinención de la lectura: desde la poética
de la obra abierta hacia la hipermedia..... 329**Nieves Pena Sueiro y Sandra Álvarez García**El *Catálogo y Biblioteca digital de relaciones de sucesos*: bases de datos
bibliográficas, textos e imágenes..... 335**Santiago Pérez Isasi y José Ramón Carriazo Ruiz**Ensayo de un índice conceptual para un corpus literario de textos
alineados multilingües 347**Mónica Poza Diéguez**The Littera Project: *Webpapers* o trabajos académicos 2.0 361**Ernesto Priani Saisó y Ana María Guzmán Olmos**

TEI como una nueva práctica de lectura..... 373

Begoña Regueiro Salgado

Innovación y creatividad: enseñar literatura en el siglo XXI..... 383

María José Rodríguez-Espiñeira y Yolanda López ArcaEl proyecto MORFOGEN. Diseño de una aplicación web para visualizar
familias léxicas 395**Dolores Romero López**Hacia la *Smartlibrary: Mnemosine*, una biblioteca digital de textos literarios
raros y olvidados de la Edad de Plata (1868-1936). Fase I..... 411**Laura Sánchez Gómez**

Poéticas Digitales: el texto como imagen de una obra cinética..... 423

**Margarita Santos Zas (coord.), Francisca Martínez Rodríguez
y Rosario Mascato Rey**

La creación y gestión del archivo digital valleincliniano: *corpus manuscrito
e impreso* 435

Sabine Thuillier

De una obra digitalizada a una obra digital:
reflexiones en torno al *Diccionario Griego-Español* en línea 459

**Gael Vaamonde, Ana Luísa Costa, Rita Marquilhas, Clara Pinto
y Fernanda Pratas**

Post Scriptum: Archivo Digital de Escritura Cotidiana 473

PRESENTACIÓN

Esta monografía reúne una selección de trabajos presentados en el I Congreso Internacional de *Humanidades Digitales Hispánicas. Sociedad Internacional (HDH)*, que se celebró en la Universidade da Coruña (UDC), de España, los días 9-12 de julio de 2013, con el título **Humanidades Digitales: desafíos, logros y perspectivas de futuro**, organizado por el SIELAE (Seminario interdisciplinar para el estudio de la Literatura Áurea Española) del Departamento de Filología Española y Latina de la UDC. Los trabajos que, una vez terminado el congreso, fueron enviados para publicación, han sido sometidos a una evaluación por pares (*peer review*), y sólo aparecen aquí los que la han superado. A su vez, todos los trabajos presentados en el congreso ya habían sufrido una evaluación por pares de un comité de 18 especialistas, como más abajo se detalla¹.

La HDH, inscrita desde agosto de 2012 en el registro oficial de asociaciones de España, surgió de la necesidad (largamente sentida por muchos especialistas y estudiantes de Humanidades) de tener un ámbito de discusión, información y reunión que nos permitiese estar al día en las tareas que combinan las metodologías de las disciplinas tradicionales de las Humanidades con herramientas proporcionadas por la informática y la estadística (codificación de datos y metadatos, visualización de los datos, procesamiento del lenguaje natural, lingüística computacional, recuperación de información, minería de textos, etc.), así como la publicación digital y ediciones filológicas con anotación y marcación interna.

Poco a poco en los últimos años se han ido interesando un número creciente de investigadores en desarrollar, probar y aplicar nuevos métodos para extraer información de la mayor cantidad y variedad de fuentes posible, y lo más interesante: plantear nuevas preguntas de investigación, brindar mejores respuestas a las antiguas preguntas, y apoyar las respuestas con muchos más datos.

Hay todavía quien piensa que no existen las Humanidades Digitales (un nombre que ha tardado en fraguar, pero que parece que ahora permanecerá en esa forma tras muchos titubeos). Esa tendencia escéptica (de antigua raigambre en España) se caracteriza por abominar de toda relación entre lo que se ha considerado el terreno sublime de las Humanidades, en especial la Filología, y las destrezas técnicas asociadas por algunos con lo propio de disciplinas auxiliares subalternas.

¹ El índice se ha estructurado encabezando los trabajos que fueron presentados como ponencias plenarias y el resto se ha organizado por orden alfabético de apellido de sus autores.

No queda tiempo para estas disquisiciones. La realidad se ha impuesto en los últimos años, y a pesar de que aún quedan muchos que, como decía el poeta Antonio Machado, *desprecian cuanto ignoran*, pertenecen sin lugar a dudas al pasado. Quienes no evolucionan perecen, y por lo tanto, no hemos de perder en tiempo en considerar sus posturas que no son sino disimulos de tu torpeza para adaptarse a un mundo en permanente y vertiginoso cambio. Tampoco nos debemos dejar obnubilar por quienes venden la etiqueta "Informática + Humanidades" como un marbete que pretende fascinar con palabras abstrusas, amparándose bajo el medio instrumental para disimular la poca sustancia de sus investigaciones. En el medio está la virtud, como decía Horacio.

La realidad es que, lo queramos o no, asistimos a un cambio de paradigma intelectual, muy semejante al que vivieron los Humanistas con la generalización del invento de Johannes Gutenberg. Durante unos 550 años ha dominado la tecnología de la imprenta, y continúa con nosotros, adaptándose con esfuerzo a los nuevos tiempos; pero en los últimos 25 años, la relación entre intelectuales y ordenadores primero y luego la generalidad del uso de Internet, ha cambiado nuestros cerebros y revolucionado las prácticas intelectuales y ha modificado nuestras costumbres y prácticas sociales.

El papel del humanista como agente social y cultural, que tanta importancia tuvo en el pasado, bien influyendo en la formación de élites y gobernantes o en la creación de una concienciación crítica en la sociedad, sólo podría mantenerse si el humanista actual es capaz de dominar las técnicas digitales y crear sus propias herramientas, sin esperar que otros no capacitados las desarrollen. Tal vez podamos achacar la pérdida de influencia de las Humanidades en el mundo actual a una excesiva parsimonia o incluso renuencia de los especialistas en estas materias a incorporar las nuevas tecnologías a las otras habilidades adquiridas en la Universidad para acometer sus investigaciones. En su favor, hay que decir que, hasta ahora, quien ha sentido interés por aunar los conocimientos humanísticos con la informática ha tenido que buscarse por sí mismo la formación, generalmente fuera de la universidad, realizando un notable esfuerzo que, afortunadamente, podemos comprobar en este volumen que ha recompensado a quienes emprendieron ese camino.

Llevamos varios años recibiendo alumnos en la universidad de los que se denominan "nacidos digitales", y de ningún modo puede la universidad seguir ignorando la necesidad de una formación adecuada en Humanidades Digitales, impartida desde sus aulas y con programas adaptados a las necesidades de las diferentes especialidades humanísticas.

La joven asociación HDH pretende ser un catalizador que avive y empuje a las instituciones académicas a valorar la investigación en este campo y a desarrollar formación en todos los niveles (grados, postgrados, doctorados) porque es lo que demanda la sociedad.

En este congreso hemos recibido 63 propuestas (algunas de ellas son paneles con más de una comunicación, o sea que en total resultaron 69). Todas ellas se inscribieron en la aplicación *Easy Chair* para ser sometidas a una rigurosa evaluación por pares (*peer review*) por un comité científico formado por 18 especialistas, que sólo aceptaron las que consideraron con un nivel adecuado.

El congreso consiguió el patrocinio de la *Alliance of Digital Humanities Organizations* (ADHO), lo que avala la seriedad de nuestros planteamientos y actuaciones. Aparte de las comunicaciones, paneles y pósteres, pudimos asistir a

cinco ponencias plenarias que aportaron una visión de la experiencia en diversos campos de las Humanidades Digitales. También se mantuvieron dos sesiones de discusiones dirigidas que resultaron de enorme interés entre los asistentes:

1. "Hacia un plan de estudios de referencia para las Humanidades Digitales", coordinada por José Luis Canet, con participación de Paul Spence y Alexandro Bia.
2. "Investigación en Humanidades en el entorno virtual: creación, difusión y valoración académica", coordinada por Nuria Rodríguez, con participación de Nieves R. Brisaboa y Domenico Fiormonte.

Creemos que el congreso sirvió para echar a andar la nueva asociación HDH, así como para conocer y dar visibilidad y difusión a los proyectos que se están llevando a cabo sobre Humanidades Digitales, especialmente los que se desarrollan en España y Portugal. Además, la reunión científica sirvió también para crear y estrechar vínculos intelectuales, profesionales y amistosos entre las 104 personas que realizaron el esfuerzo de trasladarse desde 11 países hasta A Coruña: Canadá, Colombia, Francia, Hungría, Italia, México, Portugal, España, Suiza, Reino Unido y Estados Unidos.

Agradecemos a la *Universidade da Coruña* el apoyo recibido, así como las facilidades en el uso de infraestructuras en la Facultad de Filología.

Como presidenta del *Comité local organizador* del congreso, manifiesto el agradecimiento a todos los miembros del comité, que hicieron posible la realización del encuentro: Nieves Pena Sueiro (Secretaria), Eduardo Rodríguez López, Félix Córdoba Rodríguez, Rocío Barros Roel, Carlota Fernández Travieso, Sandra Peñasco González, Andrea Maceiras Lafuente, David Loureiro González y Montserrat Muriano Rodríguez.

SAGRARIO LÓPEZ POZA
Catedrática de Literatura Española
Universidade da Coruña. España
Presidenta del Comité local organizador del Congreso

Reflexiones sobre las humanidades digitales¹

JOSÉ LUIS CANET
Universitat de València
Jose.canet@uv.es

Quiero empezar dando las gracias a Sagrario López Poza y a la Sociedad Internacional *Humanidades Digitales Hispánicas* por haberme invitado a este primer Congreso Internacional, sobre todo porque mis relaciones con la Universidad de A Coruña son muy fluidas y he participado en muchos eventos sobre tecnologías digitales en diferentes ámbitos de las humanidades, y por tanto es un placer estar de nuevo aquí y en este primer Congreso.

Cuando me invitó la profesora Sagrario López no tenía muy claro el tema de mi comunicación, y me comentó que, por mi dilatada trayectoria en algunas facetas de las Humanidades digitales, podía hacer unas pequeñas reflexiones sobre mi experiencia y los posibles cauces futuros, muy en línea con el título del Congreso: *Humanidades digitales: desafíos, logros y perspectivas de futuro*.

Empiezo pues por el pasado, o mi experiencia anterior, por si hubieron esos logros que se indican en el título del Congreso. En la década de los ochenta del siglo XX, algunos profesores de Humanidades (en mi caso de Literatura Española) nos planteamos la necesidad del uso de ordenadores como soporte a la investigación. Fue así como me inicié en la computación. Todavía recuerdo los primeros IBM, Commodore o Amstrad con sistema operativo MS-DOS y sin disco duro, sólo con *floppy disk* que incorporaba el sistema operativo. Muy pronto me compré el primer Apple. Estoy hablando de ordenadores muy caros en su momento, y muy pocos fueron los que quisieron invertir económicamente en estos aparatos (sólo indicar que el segundo Macintosh que me compré, ya con disco duro interno, me costó unas 400.000 pesetas de la época); una gran parte del profesorado no se incorporó en esta primera fase a las computadoras, pues sólo veían la posibilidad de tratamiento de textos, y para ello ya existían máquinas de escribir con memoria mucho más baratas.

Yo empecé muy pronto explorando las posibilidades de las bases de datos, sobre todo las sencillas bibliográficas. Poco a poco me interesó poder profundizar en ellas, y primero con Cuarta Dimensión (base de datos francesa) y posteriormente con FileMaker, empecé a desarrollar mi primera base de datos rela-

¹ Este trabajo se ha realizado dentro del Proyecto de Investigación del Ministerio de Economía y Competitividad, referencia FFI2011-25429.

cional (que servía como gestor completo bibliográfico y que también trasladé a otros campos de gestión universitaria). Fue así como surgió una amplia relación con Apple, incluyéndome entre sus desarrolladores a partir del año 1987, por lo que recibía todas las innovaciones de sus sistemas operativos y diferentes programas para su evaluación. Fue por esa época cuando decidí incorporar en ciertas secciones de la Universitat de València la primera red Apple Talk para la gestión, al mismo tiempo que se estaba desarrollando la red Telnet mediante los ordenadores IBM en nuestro Centro de Cálculo, todo ello funcionando como cliente/servidor.

Poco tiempo después, en el año 1990 fui nombrado Director de las Bibliotecas de la Universitat de València con la misión, por parte del Rector, de modernizar los sistemas de acceso a la información bibliográfica. Para ello entré en contacto muy estrecho con el Centro de Cálculo para el desarrollo de bases de datos de gestión bibliotecaria; finalmente decidimos, después de analizar los programas que funcionaban con el IBM central, la compra de la base de datos bibliotecaria Dobis/Libis. Un año después nació el Gopher, antecedente del Worl Wide Web, en el que la información se estructuraba en forma de árbol y mediante archivos de texto ASCII. Antes ya existía el FTP (*File Transfer Protocol*) para descarga de archivos, y con este sistema nació la primera Biblioteca virtual, el *Proyecto Gutenberg*, que posteriormente se estructuró en web. Yo ya conocía dicha biblioteca virtual y mediante el protocolo Gopher monté un primer servidor Apple con textos medievales españoles (únicamente incluí tres o cuatro ediciones más y varios artículos, claro está en texto ASCII, sin ninguna posibilidad de negritas, cursivas o marcas de cualquier tratamiento de textos). En el año 1992 empecé el diseño de la primera página web del SIB (Servicio de Información Bibliográfica de la Universitat de València), a través de la cual ya se podía acceder a parte del catálogo de nuestras bibliotecas mediante Telnet o a través de la pasarela Z39.50, pensada específicamente para bibliotecas, la cual se puso en funcionamiento en 1993 (se puede ver una muestra algo posterior en Internet Archive: <<http://web.archive.org/web/19970607060549/http://www.uv.es/~infobib/>> [29/09/2013]). También se incluía información sobre diferentes recursos en línea de textos electrónicos. Esta página web me sirvió para empezar con el portal *Lemir*, donde incorporé ya textos en Html y empecé en 1995-1996 con una de las primeras revistas cien por cien electrónicas salidas en España sobre Literatura: *Lemir, Revista de Literatura Española Medieval y del Renacimiento* (se puede ver una muestra algo posterior en Internet Archive: <<http://web.archive.org/web/19970607204833/http://www.uv.es/~lemir/>> [29/09/2013]).

La web evolucionó muy rápidamente, sobre todo los protocolos del W3C y las bases de datos. Como he comentado antes, fuimos, creo, una de las primeras universidades que incorporamos el sistema Dobis/Libis de gestión bibliotecaria en España, junto con la universidad de Deusto, puesto que cuadraba perfectamente con la DB2 de IBM implementada en el Mainframe del Centro de Cálculo. Pero a partir del desarrollo progresivo de las bases de datos distribuidas, pero sobre todo por estar en una época de crecimiento económico, y coincidiendo con que la Universitat de València iba a celebrar su 500 aniversario en el año 2000, tuve la oportunidad de dirigir uno de los primeros proyectos de Biblioteca digital de España, patrocinado por la Fundación General de la Universitat de València, IBM y el Banco de Santander, con un importe de 50.000.000 de pesetas. En dicha base de datos se incorporaron objetos, es decir, los libros digitaliza-

dos en jpeg. Aspecto que hoy prácticamente incluyen muchas de las bibliotecas universitarias y nacionales. En aquel momento nos basamos en DB2, imitando la base de datos desarrollada por la propia IBM para el museo del Hermitage de San Petersburgo². Yo mismo desarrollé en la misma época otra base de datos en FileMaker para textos antiguos, consultable en línea, donde se incluyó el fondo histórico de la Universitat, y cuya principal característica fue el diseño de la familia de letras utilizada para poder visualizar la descripción de los ejemplares tipográficamente. Posteriormente esta base de datos se transformó, mediante exportación, en diferentes volúmenes de nuestro Catálogo del siglo XVI y XVII.

Dos años antes, en 1998, me concedieron el Proyecto de investigación, subvencionado por el Ministerio de Educación y Ciencia: *Parnaseo: Servidor web de Literatura Española*, de donde nació el portal Parnaseo (<<http://parnaseo.uv.es>>), en el que se incluyeron inicialmente varias revistas digitales y diferentes bases de datos bibliográficas específicas de la Literatura Española y sobre bibliografía valenciana del siglo XVI (portal que sigue funcionando en la actualidad). También se incorporó la colección de textos medievales y renacentistas existente en el anterior portal *Lemir*, pero con la inclusión ya de JavaScript y Frames. Todo se estructuró en lenguaje HTML puro y duro, excepto las bases de datos, en un intento de que su formato tuviera la misma visualización en los diferentes navegadores y sistemas operativos. Esta ha sido siempre una de las premisas en mis desarrollos: la no utilización de aplicaciones específicas de un sistema operativo que impidiesen una visualización correcta en otros ordenadores y navegadores web, como tampoco el uso de applets, por la reticencia de muchos usuarios a descargarse cualquier elemento para visualizar una página.

En el año 1999, la Real Academia Española puso en marcha el proyecto CREA (Corpus de Referencia del Español Actual) y CORDE (Corpus Diacrónico del Español). Me encargué para el CORDE de un equipo de trabajo de entre 10 y 15 personas para la codificación en XML de gran parte de los textos literarios españoles medievales y del Siglo de Oro. Proyecto que duró hasta el año 2004.

Recientemente, he sido nombrado director del Servicio de Publicaciones de la Universitat de València en el año 2011. En este caso, mi interés ha sido la transformación de las publicaciones hacia lo digital, sin perder por ello la posibilidad de editar en papel. En estos momentos realizamos los ejemplares en diferentes formatos: PDF profesional para impresión en papel e impresión bajo demanda; ebooks en formato ePub con diferentes variantes para el Kindle de Amazon y Apple. Al mismo tiempo, hemos incluido todos los ejemplares del catálogo en Google (la mayoría en PDF y otros en ePub), donde se pueden visualizar unas cuantas páginas de forma gratuita, pero también en el Store Books de Google con visualización en streaming y mediante descarga para la venta de los ejemplares en diferentes formatos digitales, lo mismo que en nuestro portal de libros electrónicos.

Y finalmente, este mismo año he terminado una base de datos sobre *Tipobibliografía valenciana de los siglos XV y XVI* (<http://parnaseo.uv.es/tipobibliografia/Tipobibliografia.html>), creo que la única que existe sobre tipobibliografía, reali-

² J.L. Canet, "La nueva Biblioteca Digital de la Universitat de València", *JBIDI 2000: primeras Jornadas de bibliotecas digitales*, Valladolid, 6 y 7 de noviembre de 2000, pp. 71-78, <http://infonautica.net/docs/jbidi/jbidi2000/07_2000.pdf> [29/09/2013].

zada con SQL y Php, en donde se incluye la descripción de los ejemplares en PDF (con lo que se puede mantener la letrería específica diseñada para la descripción), la portada, páginas preliminares y colofón en Jpg, y se desgranar los grabados, escudos, filigranas y las letras capitulares y mayúsculas utilizados en cada ejemplar. En esta base de datos se puede buscar por autor, título de la obra, pero también por letras utilizadas por los impresores en cualquier período de tiempo, por grabados, etcétera.

Este sería un poco el pasado, en el que, relacionado con la temática de este Congreso, he participado en muchas de las líneas propuestas: edición digital de libros electrónicos; edición digital de revistas; creación, preservación y gestión de bibliotecas y archivos digitales; investigación basada en ordenadores y aplicaciones informáticas para estudios literarios, lingüísticos, culturales e históricos, incluyendo la literatura electrónica, análisis de textos, corpus, marcación de corpus, bases de datos, etcétera.

¿Qué puedo extraer de mi experiencia y de otros casos similares en nuestro país? En primer lugar el autoaprendizaje. Creo que los pioneros en las Humanidades digitales hemos tenido que ir avanzando individualmente, dedicando gran parte de nuestro tiempo al estudio de las tecnologías de la computación e intentando aplicarlas a nuestro propio ámbito de necesidades profesionales e investigaciones académicas. En segundo lugar, la necesidad, sobre todo en grandes proyectos, de trabajar en equipo, la interdisciplinariedad; lo normal ha sido incorporar a ingenieros y técnicos informáticos para el desarrollo de las propuestas. En mi caso, ha sido necesario encontrar un lenguaje comunicativo común, puesto que el licenciado en humanidades era, y creo que es, desconocedor de los términos usados en el ámbito de las nuevas tecnologías. En tercer lugar, que la evolución tecnológica avanza a velocidades impensables, mucho más rápida que en otros ámbitos del saber, por lo que estar al día de los progresos produce desánimo, al darse uno cuenta que lo que se está realizando puede estar obsoleto en un plazo muy breve de tiempo (incluso antes de terminarlo). Por tanto, se impone en este ámbito del conocimiento el aprendizaje continuo, tal y como proponen los cambios educativos de Bolonia.

Y ya que he nombrado el cambio educativo reciente en la enseñanza superior, paso al apartado docente (otro de los ámbitos de este Congreso, según el Programa con una Sesión de Discusión Dirigida, en la que participaré posteriormente). ¿Cuáles deberían ser los estudios en Humanidades digitales? O incluso, ¿deberían haber estudios específicos en Humanidades digitales? Preguntas de difícil respuesta, pues para generar un verdadero plan de estudio debemos definir la empleabilidad. Y por supuesto, como se ha indicado en diferentes foros y en comunicaciones a congresos internacionales en Humanidades digitales, existen muchas actividades diferenciadas y difícilmente se pueden englobar todas ellas en un mismo plan de estudio. Y profundizar en un aspecto concreto impide que sea un grado³ o máster generalista, por lo que queda relegado a postgrados

³ Mucho más complejo es la creación de un Grado en nuestro país, pues el Real Decreto 1393/2007, de 29 de octubre, por el que se establece la ordenación de las enseñanzas universitarias oficiales, obliga a la adscripción del correspondiente título de Graduado o Graduada a una rama de conocimiento, y en este caso sería a Artes y Humanidades o bien a Ciencias, vinculado además a un centro, lo que genera problemas serios de interdisciplinariedad a la hora de confeccionar un Plan de Estudios al intervenir los Departamentos afectados y buscar soluciones a su crecimiento profesoral; ello sin contar con que un Grado debe ser generalista,

y doctorados muy específicos. Esta es la propuesta de muchos centros en *Digital Humanities* de universidades inglesas y americanas, caso del King's College of London (*Centre for Computing in the Humanities*), University College London (UCL *Centre for Digital Humanities*), University of Oxford (*Digital.humanities@Oxford*) o University of California-Los Ángeles (*UCLA Center for Digital Humanities*).

Si analizamos los planes de estudio de algunos de estos centros, lo primero que resalta es el alto contenido en recursos computacionales. Veamos algunos casos. El máster en *Digital Humanities* de la UCL, propone los siguientes módulos:

- *Recursos Digitales en las Humanidades* (introduce a los estudiantes en una variedad de temas relacionados con el diseño, la creación, gestión y uso de los recursos electrónicos de las humanidades. Los temas incluyen recursos de humanidades en la web, la creación de recursos electrónicos, imágenes digitales, sistemas de metadatos, codificación y sistemas de marcado, diccionarios electrónicos, análisis de textos, publicación electrónica académica en las humanidades, etcétera).
- *Tecnologías de Internet* (la exploración de los conceptos básicos de marcas [XHTML], estructuración y diseño de página web; generación y entrega de contenido en línea).
- *Introducción a la programación y base de datos* (introducción a los conceptos y principios fundamentales a las bases de datos relacionales: consulta y manipulación).
- *Servidor de programación y datos estructurados* (este módulo proporciona una introducción a los temas, técnicas, tecnologías y principios subyacentes asociados con la creación y el mantenimiento de los servidores web y sitios web con bases de datos, mapas y datos espaciales).
- *XML* (proporciona una visión general del lenguaje de marcado, dando a los estudiantes la oportunidad de practicar las técnicas de marcado XML y procesamiento de XSLT).

El programa de postgrado de la UCLA –Universidad de California, Los Ángeles– prepara a los estudiantes para trabajar en el entorno digital, proporcionándoles conocimientos sobre las herramientas, métodos y problemas teóricos subyacentes de las nuevas tecnologías. Incluyen módulos sobre el análisis de textos, minería de datos, visualización, modelado y simulación, cartografía geoespacial, la narración multimedia, análisis de redes, diseño de interfaces, etcétera.

El máster *Digital Humanities* del King's College London se estructura a través de dos módulos: Métodos y Técnicas de las Humanidades Digitales y Herramientas y recursos de Humanidades Digitales. A los que se añaden una serie de cursos opcionales, que van desde la edición digital a la visualización digital en arqueología, artes, historia, cultura del libro, etc. Está concebido ante todo para desarrollar una comprensión crítica de las tecnologías digitales y la investigación en las artes y las humanidades; para aprender un conjunto de habilidades computacionales que permitan la creación de recursos digitales y que también pueda abrir nuevas e interesantes perspectivas profesionales a los estudiantes.

argumentando el interés académico, científico y profesional del mismo, avalados por otros planes de estudio nacionales o internacionales de calidad e interés contrastado (que de momento no existen en casi ningún país europeo); datos y estudios de la demanda potencial del título; etcétera.

Por último, comentaré algo sobre el máster en *Humanidades Digitales* de la Universidad Castilla La Mancha (recientemente desaparecido), subdividido en dos títulos de postgrado: *Especialista en Gestión de Contenidos Digitales* y *Especialista en Digitalización del Patrimonio Cultural*. El tronco común lo conforman los módulos: Técnicas y herramientas (18 créditos) con tres asignaturas: *Las TIC aplicadas a las Humanidades*, *Gestión de la información* y *Lenguajes de Marcado*; y el segundo módulo: *Industrias y contenidos digitales*, con dos asignaturas: *Gestión de Contenidos Digitales* y *Creación y Gestión de Industrias Culturales*. A ello se le añaden unas cuantas materias optativas que van desde la Ingeniería lingüística hasta el Periodismo Digital.

Si hay algo que es coincidente en todos estos programas de máster y/o postgrados es el alto contenido en TIC, de ahí que la mayoría de los programas estén desarrollados o dirigidos por centros o departamentos en Ingeniería computacional o Informática.

Y yo me pregunto, ¿a qué público se dirigen estos cursos de especialización en Humanidades Digitales? ¿Qué salidas profesionales plantean? Por el contenido de los programas, la mayoría piensan en graduados que procedan de las diferentes ramas de Humanidades, pero sobre todo tecnológicas, de ahí que gran parte de las materias se centren en técnicas y tecnologías de Internet (incluyendo desde gestión de servidores, hasta confección de bases de datos, lenguajes de marcas, etc.). Y prácticamente no existen materias específicas en humanidades, a no ser las que tienen que ver con desarrollos aplicados a la edición, corpus textuales, gestión del patrimonio, etc. Por tanto, pienso que las posibles salidas profesionales que están subyacentes son aquellas que tienen que ver con la investigación, como indica el preámbulo al máster del King College of London: "Su propósito fundamental es investigar cómo las metodologías digitales se pueden utilizar para mejorar la investigación en disciplinas como la Historia, Literatura, Idiomas, Historia del Arte, Música, Estudios Culturales y muchos otros". Por tanto, centrados como están en la investigación, la empleabilidad va dirigida hacia la enseñanza y a proyectos de investigación universitarios en las diferentes áreas del saber humanístico, pero muy poco hacia trabajos empresariales (exceptuando la administración pública). Otro aspecto a considerar es que difícilmente un graduado en Humanidades puede aprender en el plazo de un año todas las técnicas que se indican, a no ser alguien con un cierto grado de conocimientos en informática y computación. Y mi pregunta es, ¿no se estarán realizando estos másteres para ampliar el abanico de posibilidades de trabajo a los graduados en informática, con pequeñas técnicas aplicadas a contenidos en humanidades?

Voy a poner unos pocos ejemplos relacionados con lo que mejor conozco. En varios programas existe un módulo y a veces una orientación hacia la creación de Bibliotecas digitales. Mi primera pregunta es: ¿cuántas empresas crean bibliotecas digitales? Si son bibliotecas universitarias, de organismos autónomos o de cualquier ministerio, lo normal es que no se desarrolle ningún programa de gestión bibliotecaria digital, sino que ya exista o se compre alguno de los que compiten en el mercado. Un buena base de datos de gestión bibliotecaria necesita muchos años de desarrollo y mucho especialista implicado. Por ejemplo, en universidades muy tecnológicas que se plantearon en la década de los noventa del siglo XX el desarrollo de bases de datos para la gestión de su biblioteca, después de años de trabajo y de implementación, decidieron finalmente la compra de programas comerciales probados y experimentados. Y es que son bases

de datos muy complejas, relacionales y modulares, capaces de gestionar todo tipo de material estructurado junto con objetos. Es el caso, por ejemplo, de la base de datos *Millennium*, la más utilizada por las bibliotecas universitarias españolas en la actualidad, que funciona con Java y Oracle, mediante módulos que hacen factible la gestión completa de toda la información.

Está claro que la mayoría de las empresas no tienen la posibilidad de comprar bases de gestión tan caras por lo que podrán crear alguna para la administración de sus archivos, tanto administrativos como de venta. En dicho caso, para desarrollarla será imprescindible saber en qué sistema operativo se va a trabajar. No es lo mismo hacer desarrollos en FileMaker o Acces que en SQL, Oracle, DB2 o DB3. Por tanto, en cualquier gran empresa con Mainframe, todo se incluirá según su sistema operativo y de gestión. En dicho caso, será imprescindible un gran conocimiento del sistema y su programación, imposible de adquirir en un máster en Humanidades digitales. Para pequeñas empresas, lo más normal es que sea necesario un gran conocimiento de programación SQL, lo que tampoco es fácil de dominar en un curso de 6 ó 12 créditos.

Pongamos otro ejemplo, que es reincidente en estos estudios de Humanidades digitales: el de la digitalización del patrimonio cultural. Es un caso parecido al de gestión bibliotecaria. Para gestionar o digitalizar el patrimonio, bien sean pinturas, sellos o libros, se necesita una base de datos de gestión documental. Si se trata de la colección de una biblioteca, ya hemos comentado antes que se utilizará una base de datos modular orientada a objetos, normalmente ya existente en el organismo. Si se crea de nuevo, necesitará de un especialista en programación según el sistema utilizado. Lo importante, en dicho caso, será la creación de los campos para que sean fácilmente accesibles por el usuario final y que la información de cada objeto tenga los datos bien definidos y estructurados. Pero una vez finalizada la base de datos, una gran parte de los registros se incluirán de forma automatizada, caso de las imágenes, bien en formato JPG, bien en PDF o en otro soporte. Eso sí, para que sea funcional y válida, los campos descriptivos tienen que estar perfectamente confeccionados, de ahí que sea imprescindible que los objetos estén de antemano bien catalogados (si son libros, por ejemplo, una descripción en formato Marc). Por tanto, implica un conocimiento en catalogación, que podrían aportarlo los graduados en Biblioteconomía o en Historia del Arte, por citar los dos casos propuestos. Pero para incluir la información lo podría hacer cualquier licenciado o graduado sin necesidad de tener una formación específica en Humanidades Digitales.

Bajo estos planteamientos, la mayoría de los másteres en Humanidades digitales están pensados más para la investigación departamental o para proyectos de investigación concretos que para la empleabilidad en empresas (excepto las públicas). Por tanto, hablamos de un mercado potencial muy restringido en la actualidad. Por otra parte, la demasiada orientación informática con un alto nivel de conocimientos matemáticos espanta a los posibles estudiantes en humanidades, como resaltó la *Alliance of Digital Humanities Organizations (ADHO)* en su reunión del 2012, al analizar el fracaso de muchos de los programas que se implantaron y fracasaron en Europa de Humanidades digitales⁴.

⁴ "Digital Humanities as a University Degree: The Status quo and Beyond", <<http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/digital-humanities-as-a-university->

Para unos posibles estudios futuros de grado o de máster en Humanidades Digitales, lo primero que habría que pensar es en la empleabilidad. Si analizamos el trabajo realizado en el *Libro Blanco para el Título de Grado en Estudios en el ámbito de la lengua, literatura, cultura y civilización*, en el que participé en los años 2003-2005, se constató que el 51% de los egresados no realizaba un trabajo que correspondiera con sus estudios. Y en ello no podemos volver a caer. Si revisáramos pormenorizadamente los empleos de los estudiantes egresados en las diferentes filologías durante ese periodo de tiempo, podríamos comprobar que un 50% se dedicaba a la enseñanza, pero el otro 50% trabajaba en administraciones públicas, industria de la cultura, asesoramiento lingüístico, traducción, medios de comunicación, bibliotecarios y documentalistas, asesoramiento cultural, etcétera.

Pienso que hacia estos dos grandes grupos deberían dedicarse los egresados de los másteres en Humanidades digitales:

- a) *Enseñanza digital*. Estudios que completen la función docente en el aula, mediante el dominio de las pizarras digitales; creación de contenidos interactivos docentes; acceso a la información pertinente en Internet: manejo de catálogos de bibliotecas, portales específicos en la materia, libros electrónicos, artículos de revistas, diccionarios, enciclopedias, etc.; estructuración de la información en los motores de búsqueda como Google, Yahoo, etc.; lenguajes hipermedia: HTML, XML, PDF, etc.; publicación en la red (título del documento, cabecera, cuerpo, metadatos, estructuración de la página web); aplicaciones específicas de lectura digital y anotación colaborativa; etcétera.
- b) *Investigación y profesiones especializadas*. En este apartado cabrían las otras propuestas analizadas anteriormente en los diferentes másteres en Humanidades digitales pensados para la investigación o para profesiones específicas, caso de: creación de colecciones patrimoniales, edición de textos y libros electrónicos, periodismo digital, ingeniería lingüística, creación de diccionarios, traducción automatizada, lenguajes de marcado, desarrollo de bases de datos y servidores temáticos, evaluación de la calidad en internet, la web semántica, las redes sociales, los juegos interactivos, la edición de videos, etcétera.
- c) Pero para mí, la función primordial de un especialista en Humanidades Digitales sería la de aunar en una persona la capacidad de interrelacionar las diferentes áreas del saber tradicional con el tecnológico. Como su nombre indica, tendría que poseer un amplio bagaje humanístico, capaz de diferenciar contenidos y aplicar la rigurosidad crítica a las nuevas tecnologías.

Pienso que cada vez es más importante en la sociedad de la información y de las TIC la capacidad de coordinar y dirigir a grupos heterogéneos, bien en

[degree-the-status-quo-and-beyond/>](#) [29/09/2013]. También se indican otros factores de la desaparición o fracaso de los másteres y cursos muy especializados, como:

- Iniciar una carrera o máster con recursos insuficientes.
- Cursos que dependen de una sola persona.
- Centrarse en una rama específica, lo que limita el número de participantes.
- Perfiles poco claros, por lo que es difícil para los estudiantes ver qué oportunidades profesionales ofrece...

la investigación, bien en la empresa. Una persona que reúna las características de los humanistas del XV y XVI, los cuales eran especialistas en gramática, retórica y poética (base de la comunicación), pero también de otras muchas disciplinas, caso de las matemáticas, astronomía, música, medicina, etc., en un intento de comprender y explicar el mundo en el que vivían. El humanista actual tendría, por tanto, que conocer los nuevos lenguajes (sobre todo las diferentes variantes hipertextuales junto con las audiovisuales), las relaciones a través de las redes sociales, la gestión de la información, la creación de recursos, etc., pero sin olvidar la base esencial, que es la capacidad crítica para saber discernir aquella información válida de la que no lo es, y al mismo tiempo capaz de crear contenidos estructurados que faciliten y hagan evolucionar el saber. Todo ello con la rigurosidad con que las ciencias humanas han evolucionado hasta el presente. Por supuesto, este humanista debería conocer los entresijos de los nuevos lenguajes hipermedia y los metadatos, capaces de entrelazarse entre sí en diferentes redes mundiales, pero también para ser visibles en la maraña de la web. También ser capaz de desentrañar la parte oculta de la información (la manipulación que se hace en la web para el control de los usuarios mediante cookies, las búsquedas en Google que definen nuestro perfil para publicidad encubierta y otras casuísticas existentes en diferentes páginas web y programas de redes sociales en los que cedemos el copyright al instalar el software, etc.). Pero sobre todo, la de generador de ideas. En las grandes empresas o en la investigación universitaria es cada vez más usual el trabajo en grupo. Lo importante en la actualidad es saber coordinar a diferentes miembros procedentes de diversas áreas del conocimiento. Y pienso que los humanistas podemos tener la capacidad de esa coordinación, siempre y cuando conozcamos, aunque sea someramente, las posibilidades que nos dan las TIC. No hace falta que el Humanista digital sepa programar bases de datos específicas, pero sí su funcionamiento mediante campos, que pueden ser textuales, hipertextuales y de objetos, para poder definir las funciones que queremos obtener; la programación la dejará al especialista. Lo mismo podríamos decir de otros muchos campos, como los corpus lingüísticos, en donde lo importante es saber lo que hay que marcar y cómo queremos visualizarlo.

En fin, pienso que estamos al inicio de las Humanidades digitales, con un gran futuro si nos atenemos a lo primordial: las Humanidades, y no dejamos que sean únicamente una evolución de los departamentos y facultades de ingenierías computacionales. Lo importante, insisto, son las ideas, mientras que muchas de las aplicaciones son simplemente soportes o instrumentos que varían a velocidad de vértigo. Los humanistas tenemos que ser conscientes que la comunicación (escrita y oral) siempre ha correspondido a la Filología y las Humanidades, y ahora la hipertextual e hipermedia no se nos puede escapar de las manos, como ha ocurrido con muchas otras disciplinas. Muchas gracias.



RESUMEN

A partir de mi experiencia personal a lo largo de 30 años en edición digital de revistas y libros electrónicos; creación, preservación y gestión de bibliotecas y archivos digitales; investigación basada en ordenadores y aplicaciones informáticas para estudios literarios, lingüísticos, culturales e históricos; análisis de textos, corpus, marcación de corpus, bases

de datos, etc., llego a la conclusión de que las Humanidades digitales han avanzado mediante el autoaprendizaje y también han sido posibles grandes proyectos a través del trabajo en equipo y la interdisciplinariedad, junto con el aprendizaje continuo. Lo que me lleva a reflexionar si debería haber estudios específicos en Humanidades digitales y, si así fuera, cuáles deberían ser. Hago un breve repaso a diferentes másteres existentes en la actualidad en EE.UU., Inglaterra y España, resaltando el alto contenido en TIC, y finalmente doy mi propuesta sobre las posibles alternativas futuras de las Humanidades digitales.

Palabras clave: Pasado, presente y futuro de las Humanidades digitales, Plan de estudio.

ABSTRACT

From my personal experience over 30 years in publishing digital magazines and electronic books; in creation, preservation and management of digital libraries and archives; in research on computers and computer applications for literary, linguistic, cultural and historical studies; in textual analysis, corpus, databases, etc., I am able to conclude that the Digital Humanities have advanced through self-learning. Furthermore, major projects have been made through teamwork and interdisciplinarity, as well as through continuous learning. This leads me to wonder if there should be specific studies in Digital Humanities, and if so, which ones should they be. I give a brief overview of different Masters currently available in the U.S., England and Spain, stressing their high content in ICT. At last, I offer my proposal for possible future alternatives of Digital Humanities.

Keywords: Past, present and future Digital Humanities, Study Plan.

Musicología, informática y la vihuela en el siglo XXI

JOHN GRIFFITHS
Monash University, University of Melbourne
jagrif@me.com

En un espacio breve de poco tiempo, uno de los frutos del uso del ordenador en la investigación humanista se ha plasmado en las Humanidades Digitales, transformándose no exactamente en una disciplina, pero sí lo suficiente para que se escriba con mayúsculas. Sin lugar a dudas, lo más significativo es la creación de un enfoque y un espacio interdisciplinar en el que las metodologías y tecnologías se pueden discutir, compartir y desarrollar sin las fronteras que habitualmente limitan la interacción entre investigadores de diversas áreas. De la misma manera que se dice a menudo que la suma de las partes es mayor que el todo, el beneficio común que ofrecen las Humanidades Digitales es la capacidad de aprender, crecer y expandirse de manera exponencial, y así mismo la oportunidad para investigadores en cada área de beneficiarse directa e indirectamente de las soluciones encontradas por otros investigadores que trabajan en rincones opuestos del gran espectro humanista.

El cambio social trae inevitablemente una variedad de respuestas desde el entusiasmo más exuberante hasta la más amarga resistencia. El advenimiento del ordenador personal es muy reciente y, aunque no haya necesidad de insistir en ello en el presente contexto, vale la pena recordar lo reciente que es y lo imposible que era hace un par de décadas prever donde estaríamos ahora; y, más aún, lo difícil que es ahora anticipar adonde vamos a llegar en unos años más. Hoy, pegado a mi ordenador noche y día, me sorprende recordar que redacté mi tesis doctoral con pluma, en un mundo donde no existía la informática, solamente hace unos treinta años. No debe ser una sorpresa: solamente hemos sustituido una herramienta por otra. Seguimos, en gran parte, haciendo lo que hacíamos antes con lápiz y papel. Es más rápido, pero no es muy diferente. Nuestros antiguos ficheros ahora se han convertido en bases de datos. Sin cambios conceptuales ni filosóficos, seguimos intentando conseguir el mismo dominio completo sobre las áreas que nos ocupan, de saber todo lo que se puede saber de los temas que investigamos, y de ejercer un control sobre nuestra área de conocimiento que nos permita tratarlo con cierta seguridad, e incluso, autoridad.

En muchas ocasiones, el éxito de la informática en la transformación de una disciplina depende más de la organización interna dentro de nuestras disciplinas que de las innovaciones conceptuales o el desarrollo de nuevas aplicaciones. En un principio, depende del apoyo de las universidades a sus investigadores inno-

JOHN GRIFFITHS

«Musicología, informática y la vihuela en el siglo XXI»

Humanidades Digitales: desafíos, logros y perspectivas de futuro. Janus, Anexo 1 (2014), pp. 21-36

vadores, sobre todo en las etapas iniciales de sus programas de investigación y, posteriormente, en las fases de expansión y desarrollo. Nuestras asociaciones profesionales también pueden desempeñar un papel fundamental. Algunos de los ejemplos presentados a continuación muestran los resultados positivos que se pueden generar, ejemplificando su impacto a nivel nacional e internacional, y su papel en conseguir la aceptación de novedades informáticas que implican ciertos cambios culturales en las normas de una disciplina. Es este apoyo institucional el que aumenta la posibilidad de lograr la masa crítica necesaria para que las innovaciones se implementen.

En el epígrafe siguiente, presento unos ejemplos de proyectos musicológicos que representan también las humanidades digitales. Son proyectos orientados en parte hacia los períodos históricos de la edad media y el renacimiento que son las que maneja con mayor facilidad, también con un énfasis en áreas dedicadas a la cultura española para que sean de máxima relevancia para lectores de otras disciplinas que deseen acceder a materiales musicales. La segunda parte de este estudio se enfoca en uno de mis proyectos actuales, la creación de una base de datos centrada en la vihuela y su música, sobre todo en España durante el período aproximado 1470-1630, y su transformación para estar disponible en internet en una especie de ambiente integral que engloba la vihuela, su música y sus tañedores.

Los desafíos que enfrentan los musicólogos son, en la mayoría de los casos, los mismos que tienen que abarcar los investigadores trabajando en las demás disciplinas humanísticas. Para el historiador –en el área que sea– entre las primeras cosas que la tecnología digital permite es el acceso a materia prima en formato digital. De importancia parecida es el acceso a bibliografía secundaria, también en formato digital, con todas sus posibilidades de búsqueda y almacenamiento, siempre con el propósito de dar al investigador la capacidad de controlar y manipular todo lo que comprende su estudio. Junto con las fuentes y bibliografía, se precisan bases de datos que nos permiten catalogar, ordenar y explotar la información a nuestra disposición con la máxima velocidad y flexibilidad posible. Estas se encuentran entre las tareas principales y comunes que nos han ocupado durante este período inicial de las humanidades digitales. En cada especialidad también suele haber objetivos particulares e individuales. En la musicología, suelen ser los que atañen a la notación musical, desde su visualización en la pantalla hasta la creación de sistemas de codificación que permitirán la búsqueda y ordenación de melodías y otro material sonoro, la incorporación de código musical dentro documentos de html o xml, y la posibilidad de añadir etiquetas a información musical con la misma facilidad que en un texto normal. Además hay desafíos en la materia del sonido digital, no sólo para escuchar sino también para la clasificación, búsqueda y recuperación. Hasta cierto punto, la comercialización musical está adelantando muchas innovaciones tecnológicas que tienen que ser incorporadas entre las herramientas que se usan en la investigación actual.

LA MUSICOLOGÍA Y LAS HUMANIDADES DIGITALES

Colaboración global: RISM, RILM, RIPM, RIDM

En el ámbito de la musicología, las humanidades digitales son las beneficiarias de una larga tradición de cooperación internacional y de proyectos conjuntos

que se iniciaron con la colaboración de la Sociedad Internacional de Musicología hace más de sesenta años. El producto principal de esta colaboración son los dos recursos bibliográficos fundamentales y globales habitualmente conocidos por sus acrónimos RISM y RILM, respectivamente el *Répertoire Internationale des Sources Musicales* y el *Répertoire Internationale de Litterature Musicale*. Fundado en París en 1952, RISM se describe en su sitio web como “a multinational, non-profit joint venture which aims for comprehensive documentation of extant musical sources worldwide. These primary sources are manuscripts or printed music, writings on music theory, and libretti... housed in libraries, archives, monasteries, schools and private collections”¹. Fundado en París en 1952, RISM es el único proyecto dedicado mundialmente a la catalogación de las fuentes escritas musicales. Fue descrito en 1959 por el distinguido bibliotecario inglés Alexander Hyatt King como “uno de los proyectos más audaces de planificación a largo plazo jamás realizados respecto a las fuentes primarias de cualquier campo de las humanidades” (Hyatt King, 1959: 54).

Originalmente publicado en volúmenes impresos, RISM se está convirtiendo gradualmente en un recurso digital de acceso abierto y gratuito. En este momento, únicamente los catálogos de las fuentes musicales a partir de 1600 están disponibles en su sitio web, pero está en proceso de expansión para cubrir la totalidad de su material. El catálogo en línea contiene más de 850.000 registros, principalmente de fuentes manuscritas, y cuenta con obras de algunos 27.000 compositores conservados en bibliotecas y archivos de 36 países.

Con sede en Alemania, RISM recibe financiación a través de varios socios alemanes e internacionales. RILM, por el contrario, tiene su sede en Nueva York y se encarga de reunir y catalogar todo tipo de publicaciones musicológicas. RILM fue una iniciativa de Barry Brook en Nueva York en 1964 y comenzó a funcionar en 1968 como complemento a RISM. RILM fue publicado en papel hasta el 2008, pero ahora está disponible exclusivamente en línea en <<http://www.rilm.org>>. Su transición al formato digital ha sido extraordinaria, no solamente debido al sistema completo de catalogación, sino también por los casi cincuenta años de material que han catalogado. En la actualidad (octubre de 2013), cuenta con 742.300 registros en línea. Desafortunadamente no es una herramienta gratuita, sino accesible a través de las distribuidoras EBSCO y ProQuest. Debido a su comercialización y a las suscripciones de las principales bibliotecas mundiales, RILM es ahora una organización completamente sostenible que no tiene que depender de financiación externa. Es una de las herramientas más importantes en la investigación musicológica y sus índices incluyen publicaciones de todo tipo: libros, capítulos de libro, artículos de revistas, ediciones musicales, tesis, y otros tipos de publicaciones.

Las iniciativas RISM y RILM se han ampliado en los últimos años por la adición de RIPM <www.ripm.org> y RIdIM <<http://www.ridim.org>>. RIPM, el *Répertoire International de la Presse Musicale*, fue fundado en 1980 y está disponible en línea por suscripción. Es un recurso de rigor para los investigadores de cualquier tema biográfico o de historia social, sobre todo los que se dedican al siglo XIX. El objetivo de RIPM es hacer accesible los primeros escritos sobre la música que aparecieron en la prensa popular a partir de mediados del siglo XVIII hasta justo antes de la llegada de la informática. Se compone de tres secciones principales:

¹ <<http://www.rism.info/en/organisation.html>> [5/12/2013].

un índice retrospectivo de revistas de música de 1759 a 1962, un archivo *online* de periódicos de música Música (Full-Text), y una biblioteca electrónica de periódicos musicales. Una vez más, este es material de extraordinaria importancia, y especialmente relevante para todos los aspectos de la sociología de la música, historia social y la historia de la recepción. Cualquiera que se haya aventurado, por ejemplo, a investigar la música en la prensa popular española en el siglo XIX no puede evitar la impresión de erudición que refleja.

RiDIM complementa los índices anteriores como *Répertoire Internationale d'Iconographie de la Musique*. Se trata de un catálogo de obras de arte que, a lo largo de la historia, representa los instrumentos musicales y la vida musical. La evidencia pictórica es una parte importante del material que manejamos en la investigación musicológica, sobre todo en la medida que avanzamos hacia atrás en el tiempo. Hay que recordar que prácticamente no sobreviven instrumentos musicales en Europa anteriores a 1500, y las representaciones iconográficas son así nuestro testimonio más fiable a su construcción y su uso. Del mismo modo, formatos de notación antiguos contienen sustancialmente menos información que las fuentes posteriores en cuanto a su interpretación y, por lo tanto, las imágenes pictóricas proporcionan una dimensión esencial de nuestro conocimiento. La clasificación de la iconografía es frecuentemente laboriosa ya que las fuentes no suelen indicar los nombres de los instrumentos y, aunque parezca una tontería, frecuentemente es una tarea ardua el conectar imágenes iconográficas con nuestras fuentes textuales. La base de datos de imágenes está en línea en <<http://db.ridim.org>>

En España, hay otras instituciones que participan en el desarrollo de recursos en línea relacionados con la iconografía musical, como el Grupo Complutense de Iconografía dirigido por Cristina Bordas <www.imagenesmusica.es> y, más recientemente, la Asociación Española de Documentación Musical <www.aedom.org> que en los últimos meses ha inaugurado un nuevo catálogo en la red sobre la iconografía musical en España, el *Catálogo de Iconografía Musical* <<https://iconografia-aedom.org>>.

PROYECTOS Y DESAFÍOS

Más allá de los cuatro proyectos excepcionales colaborativos internacionales que acabamos de comentar, hay numerosos proyectos desarrollados por pequeños equipos o investigadores particulares que responden a desafíos más generales y que proporcionan recursos que sirven a toda una comunidad de investigadores por todo el planeta. Muchos comparten fases de evolución parecidas, por ejemplo, bases de datos de proyectos individuales que acaban convirtiéndose en herramientas para un grupo extendido de expertos con intereses superpuestos, o programas iniciados por pequeños equipos de investigadores, profesores, alumnos y aficionados.

Teoría musical

Una de las áreas de investigación musicológica casi exclusivamente basada en fuentes textuales es la de la teoría de la música, un campo enorme en el que tenemos no sólo un cuerpo sustancial de los escritos de la Antigüedad clásica, sino también un gran corpus de teoría musical que se extiende desde la temprana Edad Media hasta la Era Moderna. Incluso desde el momento en que

las lenguas vernáculas se convirtieron en el vehículo del discurso académico, gran parte de los escritos de teoría de la música continuaron siendo compuestos en latín. Uno de los proyectos pioneros en la musicología digital es el proyecto TML, el Thesaurus Musicarum Latinarum iniciado en 1990 y ahora asentado en la Universidad de Indiana <www.chmtl.indiana.edu/tml>. Después de más de veinte años de continuado crecimiento, contiene una enorme cantidad de textos latinos de teoría de la música del extenso período 900-1700. Aunque todavía un recurso poco sofisticado en su formato gráfico, tiene ejemplos en la incorporación de imágenes de notación musical dentro del texto, sigue ofreciendo acceso a un número significativo de fuentes y materiales.

Canto litúrgico

Otra área enorme de la historia musical de la que los recursos digitales ofrecen un potencial sin precedentes es el estudio del canto llano litúrgico. Aquellos entregados sacerdotes que practicaban y estudiaban el canto y guardaban tanta información en su memoria casi pertenecen a un pasado irrecuperable. Bajo estas circunstancias, la informática puede jugar un papel singular en preservar una de las más antiguas tradiciones de nuestra civilización. Las oportunidades, sin embargo, se extienden aún mucho más allá y tienen el potencial de poder llegar a abarcar la enorme variedad en el repertorio que se debe a las diferencias regionales en la liturgia y el canto llano, todo lo contrario de lo que han sido los intentos de la iglesia católica a través de los siglos de unificar, estandarizar y homogeneizar la liturgia.

Los variantes regionales en canto llano se producen en las melodías tanto como en los textos, a veces en ambos a la vez. En términos de la informática, los textos presentan pocos problemas; las dificultades se han presentado en cuanto a la creación de un lenguaje normalizado para representar el contenido musical, simplemente en términos de codificación sin tener en cuenta las dificultades para su visualización y representación gráfica. Como resultado, las primeras bases de datos de canto litúrgico en internet se dirigían exclusivamente al texto. Entre las primeras se encuentra la base de datos *Cantus*. Todavía sigue disponible en su formato original en <<http://publish.uwo.ca/~cantus>>, y aunque desde mayo de 2012 no se ha actualizado, alberga más de 380.000 registros. Originalmente creada por Ruth Steiner, es un ejemplo paradigmático de proyecto particular que se une en una empresa colectiva dirigida por un grupo de expertos en EE.UU. y Canadá. Víctima de su temprana creación, su limitación principal es la ausencia completa de información musical.

Afortunadamente, *Cantus* está entrando en una nueva fase de desarrollo, a partir de su migración para formar parte del grupo de investigación en humanidades digitales MARGOT en la Universidad de Waterloo (Ontario, Canadá). Con su nueva dirección <<http://margot.uwaterloo.ca/projects/cantus/>>, *Cantus* sigue avanzando, ahora bajo la dirección de Debra Lacoste (Universidad de Waterloo) y Jan Kolářek (Charles University, Praga). La nueva versión está incorporando imágenes en facsímil para ofrecer algunas posibilidades musicales, un posible primer paso hacia una herramienta de mayor utilidad musical.

La primera base de datos del canto litúrgico en incorporar material musical como parte integral fue la *Medieval Music Database* (MMDB) <<http://www.lib.latrobe.edu.au/MMDB/>> desarrollado por John Stinson en la Universidad de La Trobe en Melbourne. Parte del proyecto inicial fue la creación de un lenguaje

informático para la codificación digital de canto llano y su visualización en pantalla. Fue publicado en 1994 y contiene unos 70.000 cantos que representan el año litúrgico completo según el uso dominico. No se ha actualizado desde 2004 debido a problemas de financiación, no obstante, sigue recibiendo un número elevado de visitas debido a su fiabilidad y la facilidad de su manejo.

Con la publicación en internet de la *Global Chant Database* <<http://www.globalchant.org>> en 2009, se inicia una nueva etapa en las herramientas digitales dedicadas al canto llano. Invento del investigador e informático checo, Jan Kolářek, ahora participando en el futuro desarrollo de *Cantus*, es el recurso más sencillo pero más sofisticado para búsquedas musicales. Igual a MMDB, las búsquedas muestran no solamente el texto litúrgico sino también el íncipit musical. Las herramientas de búsqueda agrupan los resultados en categorías de 1) coincidencias exactas, 2) transportaciones de la misma serie interválica, y 3) ocurrencias del fragmento melódico especificado dentro de las melodías, es decir, en posiciones internas y no limitado a los comienzos de los cantos.

En España, una de las innovaciones recientes es la incorporación en la web de la Biblioteca Nacional de una nueva base de datos, *Cantoriales* <www.bne.es/es/Catalogos/Cantoriales>, que pone a disposición de los investigadores la amplia colección de casi cien manuscritos litúrgicos que acabaron en dicha biblioteca después de las desamortizaciones decimonónicas. En cuanto a búsquedas, emplea la misma tecnología desarrollada para la Global Chant Database que permite tanto búsquedas melódicas como textuales. Es una herramienta de gran utilidad para la investigación española y representa un ejemplo modélico para el futuro.

Otros repertorios y áreas

Iniciativas similares en catalogación y compilación de bases de datos se están proliferando en diversas áreas musicológicas referentes a la música polifónica. Una de las primeras en desarrollarse ha sido la base de datos de la música del siglo XIV que forma parte de *MMDB* (mencionada anteriormente), que proporciona el repertorio completo organizado por autor o por manuscrito, con todas las concordancias y con una bibliografía específica para cada obra. Ejemplifica las bases de datos de repertorios determinados como la dedicada a la *chanson* francesa del s. XVI realizada por Annie Coeurdevey en el CESR en Tours <<http://ricercar.cesr.univ-tours.fr/3-programmes/basechanson/index.htm>> Comprendiendo 9711 registros, esta herramienta indica el inicio poético y musical de cada *chanson*, junto con los detalles de los autores de la poesía y la música, y las fuentes donde se conserva. Existen herramientas parecidas para otros repertorios renacentistas como el motete. La más antigua de las dos existentes es una reliquia de una primera generación de intentos en las humanidades digitales. *The Latin Motet: Indexes to Printed Collections, 1500-1600* <<http://library2.binghamton.edu/cgi-bin/lincoln/latin-motet.cgi>> es una versión en html de libro del mismo título de Harry Lincoln editado en 1993. A pesar de su falta de sofisticación, sirve para recordarnos las ventajas de la sencillez. La más reciente *Motet Database Catalogue Online* <<http://www.arts.ufl.edu/motet/default.asp>> ha estado en la red desde 2003, y es otro ejemplo de una herramienta inicialmente creada como parte de los estudios doctorales de su creadora, Jennifer Thomas de la Universidad de Florida. Incluye más de 50.000 registros para los motetes del período 1475-1600, y cuenta con la colaboración

de un destacado equipo internacional. Estas bases de datos no permiten búsquedas melódicas en este momento debido a la enorme tarea de digitalizar su repertorio y a la falta de un lenguaje normalizado musical.

Otros proyectos musicológicos se extienden más allá de los confines de los repertorios y fuentes. Un ejemplo de tal proyecto versa sobre la *Prosopographie des Chantres de la Renaissance* <<http://ricercar.cesr.univ-tours.fr/3-programmes/PCR/>> bajo la dirección de David Fialla y Philippe Vendrix del CESR de Tours. El propósito es el de crear una base de datos formada por documentos de archivos principalmente eclesiásticos y cortesanos que permitirá trazar la carrera de cantores y compositores del período 1350-1600. La versión en este momento disponible en internet todavía emplea FileMaker Pro pero próximamente será reemplazada por una nueva versión más flexible escrita en xml lo cual mejorará la interfaz del usuario. La presente versión publicada solamente emplea una proporción limitada de toda la información que alberga.

Recursos y futuros

Como hemos indicado y en común con las demás humanidades, una gran proporción de la energía dedicada a la creación de recursos digitales se ha centrado en la catalogación y digitalización de fuentes, y en proyectos bibliográficos. La música se incluye dentro de algunos de los grandes proyectos interdisciplinarios como *Gallica* de la Bibliothèque National de France y el programa de digitalización de fuentes antiguas de la Bayerische Staatsbibliothek. Más recientemente la British Library ha digitalizado unos 320 libros para su proyecto *Early Music Online* que se realiza en asociación con Royal Holloway de la Universidad de Londres <www.earlymusiconline.org>. De acuerdo con las tendencias internacionales actuales en la política de bibliotecas, estos proyectos ofrecen sus materiales de forma gratuita, reconociendo su papel en la custodia de Patrimonio de la Humanidad y sin ánimo de lucro. También entre los proyectos principales dedicados a fuentes musicales de épocas pasadas se encuentra el ambicioso *Image Archive of Medieval Music* (DIAMM) <<http://www.diamm.ac.uk>> que se ha convertido en un punto de referencia para el estudio de la música medieval. DIAMM alberga más de 15.000 imágenes de manuscritos musicales medievales junto con datos detallados de catalogación de miles de manuscritos, así como ediciones musicales que se publican gratuitamente junto con importantes tesis doctorales sobre música medieval. Para los manuscritos anteriores a 1600, sus registros están mucho más actualizado que los de RISM. Es un proyecto de horizontes abiertos y cuenta con un gran futuro.

Las fronteras de las humanidades digitales y la música, desde mi punto de vista, tienen más que ver con el desarrollo de la tecnología informática en sí y no tanto con la aplicación de tecnologías existentes a nuevos campos musicales. Veo más la necesidad de refinar los sistemas para la codificación de información musical, como hemos comentado anteriormente, y de desarrollar procesos de reconocimiento óptico de música, no solamente desde notación moderna, sino desde las distintas formas de notación original. Durante las últimas décadas se ha utilizado MIDI en este sentido, aunque últimamente vemos el nuevo formato de Music-XML cada vez con más frecuencia. Según los expertos ciertas desventajas no van a facilitar la adopción universal de este lenguaje, y hay un fuerte movimiento hacia el uso del *Music Encoding Initiative* (MEI) <<http://music-encoding.org>> –un protocolo paralelo al conocido *Text Encoding Initiative*

(TEI)- para uso integrado en XML con la capacidad de mostrar figuras musicales dentro de textos y de aplicar un sistema de marcado similar al TEI.

LA VIHUELA EN EL SIGLO XXI²

Mi proyecto principal en este momento relacionado con las humanidades digitales es la creación de un sitio integral sobre la vihuela, el instrumento que ha sido desde hace años el principal enfoque de mi trabajo. El proyecto intenta reunir todo el material que he ido recopilando durante las últimas décadas, y que he ido metiendo en bases de datos desde hace veinticinco años. La tarea ahora es de ampliarlas, completarlas, consolidarlas y conectarlas en una sola base de datos integrada. Es un proyecto que concibo como un tipo de hábitat, un entorno auto contenido y completo en sí, al mismo tiempo que una herramienta para futura investigación. Quizás es posible contemplar el proyecto de esta manera ya que es un campo reducido, capaz de ser contenido con relativa facilidad en un ambiente controlado. Se conservan solamente unas cuatro o cinco vihuelas originales renacentistas y un repertorio de aproximadamente unas setecientas obras musicales, más una variedad de material iconográfico y documental. Tal vez estas condiciones son las que permiten contemplar un entorno digital diferente a lo habitual, la posibilidad de reunir todo relativo a la vihuela bajo un solo techo, de forma íntegra. Podría describirse como una enciclopedia no narrativa del instrumento y su música o, simplemente, un microcosmos de la vihuela.

Las principales áreas de la base de datos son un instrumento en sí, la música, y las personas que las construían, las poseían, o las tocaban, durante el período aproximadamente entre 1470-1630. Estos límites cronológicos se extienden desde los comienzos del uso de la vihuela como instrumento pulsado, hasta la transformación en el siglo XVII en un instrumento de cinco órdenes de cuerdas que solía llamarse guitarra. Visto desde esta perspectiva, la vihuela es un solo instrumento que, a través del tiempo, sufre una serie de modificaciones según las necesidades de sus usuarios y en respuesta a los cambios sociales de modas y gustos musicales. Durante la Edad Media, era un instrumento pequeño, sostenido en el brazo, y tocado con un arco. Durante la segunda mitad del siglo XV se separó en dos modelos diferentes llamados vihuela de arco y vihuela de mano. Esta última, la que aquí nos interesa, fue utilizada principalmente para música de la tradición polifónica a lo largo del siglo XVI, hasta transformarse de nuevo durante la primera mitad del siglo XVII en guitarra, ahora empleada para un nuevo estilo popularizante, y principalmente tocada por rasgueado. La vihuela de mano renacentista se tocaba en contextos cortesanos y urbanas de toda España, Portugal, las colonias americanas, y las zonas de Italia bajo el control español, sobre todo Nápoles, los Estados Pontificios y Milán. Fue interpretada por músicos profesionales y aficionados, utilizada en espectáculos colectivos, especialmente en la corte, y también tocada en solitario como una actividad privada y personal, conectada con la reflexión y la vida espiritual interior.

Aparte de la información sobre los instrumentos, música y tañedores, la base de datos también está estructurada con otra dimensión documental, información referente a los documentos originales, escritos modernos, y grabaciones. Esta es-

² El presente proyecto se está desarrollando en la Universidad de Melbourne subvencionado por el *Australian Research Council*, la entidad nacional de investigación y desarrollo.

estructura –con bases independientes para documentos, bibliografía y discografía– garantiza una cierta pureza de la información contenida en ella. La tabla que registra las fuentes y los documentos originales –bien sea con texto completo, abreviado o simplemente un resumen– representa en sí el cuerpo completo de documentos en que se basa nuestro conocimiento contemporáneo de la vihuela y su entorno. En las fases iniciales de la base de datos, esta información se incluía directamente en registros individuales de diferentes instrumentos o personas. En la medida que se iba aumentando, se hizo patente que este sistema suponía demasiada repetición de información y decidimos revisar este aspecto de la estructura. En su estructura actual, la base de datos completa está formada por un conjunto de siete bases independientes, tres de los objetos de interés –instrumentos, música, personas– junto con tres que albergan las fuentes de la información –documentos, bibliografía, discografía– a las que se añade la séptima base de temas musicales que complementa la información musical, pero que merecía ser una base de datos en sí misma, o dentro del software FileMaker Pro que hemos utilizado como soporte, una tabla independiente. La estructura de la base de datos –es decir, el conjunto de tablas– se ve en la figura 1 con las flechas indicando los enlaces relacionales entre las diferentes tablas.

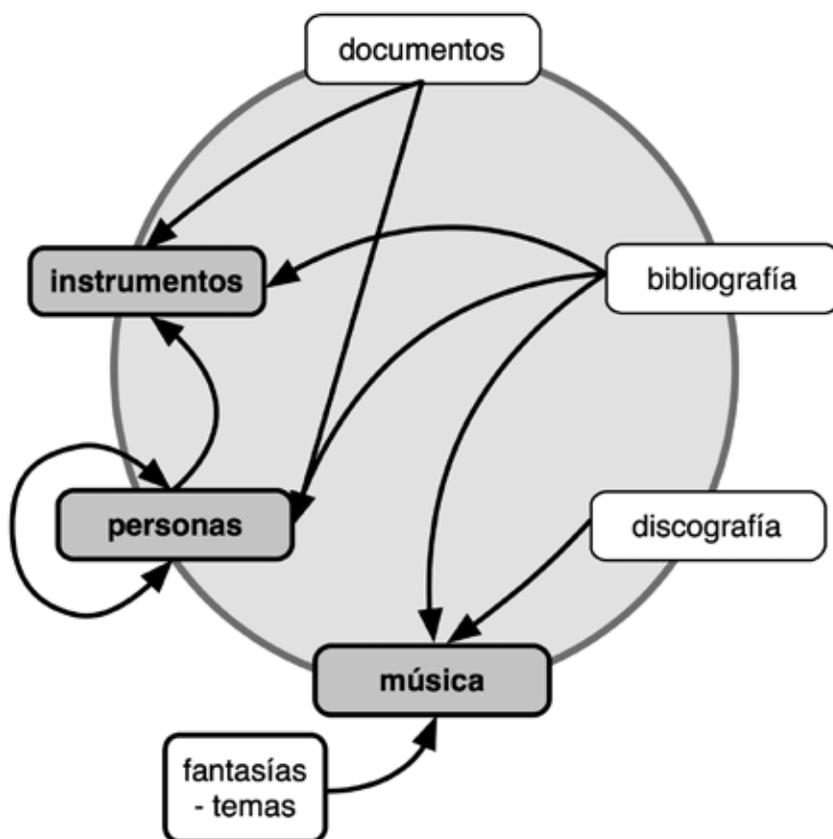


Fig. 1. Estructura de la Vihuela Database.

La base de datos es el producto de más de veinte años de recopilación. Desde el principio ha sido desarrollada con FileMaker Pro y las fases de su evolución en gran medida coinciden con los avances en el software mismo durante este tiempo. Al principio, eran cuatro bases de datos independientes: instrumentos, tañedores y constructores, música y bibliografía. Cuando FileMaker se convirtió en software relacional, fue posible integrarlas en el mismo fichero, como tablas independientes de una sola base de datos. Poco a poco se fueron añadiendo las capacidades relacionales, hasta llegar a su estado actual. Debido al proceso de acrecentamiento, fue necesario en un momento dado reconstituir el fichero de nuevo, desde cero, y según la lógica del programa actual. Esta revisión suponía la eliminación de ciertas rutinas y procesos que habían sido superados en la creación de versiones sucesivas del programa.

Para su publicación en internet, nuestra intención es de migrar los datos a un sistema diseñado con xml expresamente para la red. Hasta el momento no estoy convencido de que el interfaz de FileMaker Pro haya adquirido un nivel suficientemente atractivo o fluido para lo que se requiere en internet. La última versión del software –FileMaker 13– acaba de salir pero no he tenido la oportunidad de someterla a prueba. Por esta razón, el aspecto visual de los ejemplos siguientes no se asemeja al estilo de diseño visual que se anticipa para su publicación en internet. En primer lugar, un sitio web de este tipo deber atender a una diversidad de usuarios, desde investigadores profesionales hasta internautas aficionados. El objetivo será el de proporcionar una variedad de pantallas con diferentes niveles de información para responder mejor a las diversas necesidades de los usuarios. Aunque las estadísticas de visitas a mis páginas web indican que más del sesenta por ciento de las visitas vienen de países de habla hispana, he decidido que será preferible publicar la base de datos en inglés sin construir un sitio paralelo en castellano. Obviamente los documentos originales se mantendrán en su idioma original, lo cual de entrada dará la sensación de familiaridad al usuario hispanohablante, pero la razón principal para esta decisión se basa en crear la mayor accesibilidad internacional posible, y la convicción de que el tipo de traducción instantánea que ya se ofrece en buscadores como Google Chrome se convertirán en la norma dentro de poco tiempo.

En los próximos párrafos explicaré en más detalle cada una de las tablas que constituyen la base de datos. Todo lo que se refiere a los instrumentos está concentrado en la tabla denominada *Vihuelas* (Fig. 2). Los campos fundamentales son los que clasifican la datación y procedencia de la información, junto con el nombre del creador y la localización actual de cada artefacto, con un campo independiente donde se especifica si es instrumento, imagen, o documento escrito. En ésta y todas las demás tablas, empleo dos sistemas paralelos para la datación de los artefactos, documentos y personas para compensar la frecuente imposibilidad de atribuirles fechas exactas. Además del campo donde se registra la fecha exacta [“Date”] si se conoce, la otra es para clasificar el material por siglo, o tercios de siglo [“Century”]. Sirve para el material sin datación precisa y también para el material con fecha exacta, para poder ordenar y manipular el material fácilmente en bloques. Este sistema permite, por ejemplo, buscar con facilidad todas las vihuelas del siglo XVI en un bloque (seleccionando “16cent” en el menu del campo “Century”), o las representaciones en pintura (seleccionando “Artwork-painting” en el campo “Medium”) de vihuelas del primer tercio del siglo (seleccionando “16cent-1” en el campo “Century”). De forma complementaria, la tabla también incluye campos descriptivos donde se puede

añadir texto perteneciente al artefacto y su localización ["Artwork/Document"], o cualquier comentario sobre los instrumentos en cuestión, bien sea histórico o descriptivo ["Description"].

The screenshot displays the 'VIHUELA DATABASE' interface. At the top, there are navigation tabs for 'VIHUELAS', 'MUSIC', 'FANTASIA', 'DOCUMENTS', 'PERSONS', 'BIBLIOGRAPHY', and 'RECORDINGS'. Below these, a search bar and a table of instrument records are visible. The selected record is for a 'Vihuela de mano' from the 16th century, located in Ubeda, Jaén, Andalucía, Spain. The 'ARTWORK/DOCUMENTS' tab is active, showing a detailed entry for an artwork by Elienne Charret, circa 1515, depicting a vihuela player. The entry includes a description in Spanish and English, and a table of 'ORGANOLOGICAL CHARACTERISTICS' with dropdown menus for various parts of the instrument like the body, bridge, soundhole, strings, frets, and pegbox. A small image of the artwork is shown on the right. At the bottom, there is a 'Reproductions' table listing bibliographic references.

Instrument UNIFORM	Century	Date	Medium	City	Province (modern)	Old Kingdom	Locator	Number
Vihuela de mano	16cent/2mid	1541-1544	Artwork: Bas-relief	Ubeda (Jaén)	Andalucía	Castile	Mylla	16-218

Reproductions	Bibli. code	Bibli. content	Page	Comment
▶	gonzalez2005	González, Carlos. "La 'Vihuela virtual' de la Capilla de El Salvador de Úbeda". <i>Hispanica Lyra</i> 1 (2005): suplemento iconográfico.	n.p.	
▶	gâteau2004	Gâteau, Florence. "L'iconographie de la vihuela". <i>Actes originaux de la guitare: la vihuela de mano</i> . Ed. Joël Dugot. Paris: Musée de la Musique, 2004. 41-49.	p. 47, ill. 22	

Fig. 2. Vihuela Database, *tabla* Instruments.

Lo que ha sido de gran utilidad es la parte de la tabla en la que se clasifican las características físicas de cada instrumento [Organological Characteristics]. Mediante una serie de menús con valores preestablecidos, están recopilados detalles de la caja de resonancia, el diapason, el clavijero, las clavijas, el encordado y, en el caso de representaciones iconográficas, la forma de tocarse. Los campos referentes a la caja de resonancia ["body"] permiten clasificar el perfil de la tapa armónica, la naturaleza del fondo, detalles del puente y la(s) apertura(s) sonoras o rosetas. De las otras partes, clasificamos el largo del diapason, el número de trastes, y la forma del clavijero y desde donde se insertan las clavijas. Entre las opciones para describir la técnica instrumental, se distingue no solamente entre instrumentos pulsados y de arco, sino también la colocación en el brazo o rodilla de los instrumentos de arco, y la posición de la mano derecha en las vihuelas de mano. Son las características más significativas para la investigación del desarrollo del instrumento y de la evolución de su técnica.

Integrada en cada ficha es la bibliografía pertinente a las reproducciones iconográficas y los estudios organológicos, incorporada directamente desde la tabla bibliográfica. Los detalles de los constructores o creadores de las representaciones iconográficas también están localizados en la tabla personas y conectado de forma relacional. En este momento la tabla "Vihuelas" contiene más de 750 registros aunque solamente un sesenta por ciento son estrictamente del período principal que comprende la base de datos, de aproximadamente 1470 a 1630.

Como principio inicial, me ha parecido útil mantener los criterios amplios e inclusivos, sobre todo en las áreas de más divergencia entre opinión experta, y reconociendo que quizás más adelante algunos registros serán excluidos. Así, en este momento se incluyen un número de imágenes de vihuelas de brazo de los siglos XIII al XV, descartadas y excluidas por muchos musicólogos por no ser “vihuelas” sino “fidulas”, a pesar de ser éste un término inexistente en la lengua castellana, por lo menos en el siglo XV (Griffiths, 2010). Incluso, dentro de nuestro período cronológico principal, no se han excluido las guitarras, debido a la dificultad de distinción categórica entre ellas y las vihuelas. También figuran algunos laúdes, no solamente porque el instrumento se tocaba en España durante el período, sino porque hay evidencias en creer que el término vihuela se utilizaba en ocasiones con un sentido más genérico, como ocurre en el tratado sobre instrumentos musicales de Juan Bermudo de 1555 dónde se describe el laúd como la “vihuela de Flandes” (Bermudo, 1555: fol. 96). Por último, se incluye una selección de representantes importantes de la vihuela moderna.

El repertorio musical integral está incluido en la sección de la base de datos titulado *Music*, con un registro para cada obra. Los registros contienen información básica de las obras: título, compositor, incipit musical, fuente y, en el caso de los arreglos de polifonía vocal –el género más prolífico del repertorio– el nombre del autor de la obra original. Otros campos descriptivos indican el género musical de cada pieza, el tipo de notación empleada, el modo o tono y, en el caso de las obras para canto y vihuela, el idioma de la letra y el formato notacional de la voz cantada. La pantalla principal también incluye un campo para cualquier otra observación relevante o resúmenes de la literatura secundaria (Fig. 3).

The screenshot displays the 'VIHUELA DATABASE' interface. At the top, there are navigation tabs for 'VIHUELAS', 'MUSIC', 'FANTASIA', 'DOCUMENTS', 'PERSONS', 'BIBLIOGRAPHY', and 'RECORDINGS'. Below these, a search bar and navigation controls are visible. The main content area shows the entry for 'Soneto 30. Dichosa fue mi ventura' by Valeriano, Enriquez de. The interface includes fields for 'commentary', 'editions', 'lute tablature', 'recordings', 'analysis', 'song text', and 'retailers'. A large text box contains detailed information about the piece, including its origin (from Page Rex to 'vihuela' in July 2006) and a transcription of the sonnet's text. On the right side, there are sections for 'SPECIFICATIONS' and 'SONG FORMS', along with a musical score visualization.

Fig. 3. Vihuela Database, *tabla Music*.

Las demás pantallas de la tabla *Music* muestran la información bibliográfica en cuanto a las ediciones modernas, las grabaciones, las fuentes de la polifonía vocal adaptada para la vihuela, junto con los estudios modernos relevantes a cada obra. Este conjunto de información se suministra automáticamente desde

las tablas bibliográficas y discográficas relacionadas. Hay campos adicionales, por el momento también en pantallas diferentes, para los textos poéticos de las canciones, tanto como observaciones analíticas verbales y diagramáticas (Fig. 4).

VIHUELA DATABASE base de datos
© John Griffiths 1992-2013

VIHUELAS	MUSIC	FANTASIA	DOCUMENTS	PERSONS	BIBLIOGRAPHY	RECORDINGS
list	features sites incipits analysis	themes	list docs	list persons	biblog list	print recordings list

Daza, Fantasia 11 por el séptimo tono | Daza, Esteban | Daza | da011 | El Parrasio | 15v | Fantasia

category abstract cantus firmus dance gloss intabulation song variations miscellaneous

COMMENTARY | EDITIONS | LITERATURE | RECORDINGS | ANALYSIS | SONG TEXT | INTABULATIONS

uniform title: Fantasia 11 por el séptimo tono
source title: Fantasia por el septimo tono a tres, señalase la clauve de F#fa#ut quinta en tercero traste.

fantasia category: 1stP 1stM 1stS Par n1m n1m+1m 1d 1d+1m 1d+n1m

analysis text

Polythematic ricercar. Typical of Daza but with its own idiosyncrasies: principally a short work of 37 bars divided into 7 short sections with one theme each, of 4-7 bars (average = 5). Considerably strong relationships between themes. Themes 1-3 outline a falling 4th, 4-5 the latter is an abbreviation of the other. 6 based on 5. 7 share the falling 5th of 4. With the exception of V no section has more than a single exposition of thematic material. Voice entries are varied and give no consistent patterns. Two cadences are exceptional in that the leading tones don't resolve but have crotchet rest followed by new entries at 22-23 & 31-32. Ascend 3rd which is part of 4 5 6 has harmonic (cadential) implications which hold the work together even when theme is absent. Harmonic rhythm is the central cohesive force of the work (investigate).

Fig. 4. Vihuela Database, tabla "Music", pantalla "Analysis".

Otra característica independiente y de índole musical es el catálogo completo de los 1.180 temas melódicos empleados en las 219 fantasías originales de los vihuelistas. La fantasía era para el vihuelista el punto culminante de su arte, y el género donde cada autor mostraba su máxima capacidad como creador. En el ejemplo siguiente (Fig. 5), se muestra los siete temas empleados por Esteban Daza en análisis de la *Fantasia por el séptimo tono* que se ve en la ilustración anterior. La base de datos melódica permite búsquedas de todo su contenido por secuencias de intervalos. Esta herramienta pone de manifiesto la amplia gama de usos que puede proporcionar esta parte de la base de datos. Facilita el rápido manejo de varias de las cualidades musicales del repertorio por usuarios, independientemente de si sea un profesional, o un aficionado en busca de grabaciones y ediciones para tocar.

La tabla "Persons" alberga información sobre una amplia gama de personas, la mayoría de las cuales estaban relacionadas con la vihuela de alguna manera. Además de los tañedores, compositores y violeros, el elenco incluye a los artistas y escritores que documentaron vihuelas y cuyos obras figuran en la tabla de instrumentos. Se incluyen también los mecenas, los tañedores ficticios en las obras literarias contemporáneas, así como ciertos individuos sin conexiones musicales pero enlazados de alguna forma a vihuelistas conocidos. En la actua-

VIHUELA		VIHUELAS		MUSIC		FANTASIA	DOCUMENTS	PERSONS	BIBLIOGRAPHY	RECORDINGS
DATABASE		list	features	titles	incipits	analysis	list docs	list persons	bibliog list	print
base de datos						THEMES				recordings list
© john griffiths 1992-2013										
intervals	composer	work	theme code	opus	theme				sequence	uniform title
g -2 1 -4	Daza	Da011	Da 11.1						349	Fantasia 11 por el séptimo tono
d 2 1 -2 -3 -4	Daza	Da011	Da 11.2						642	Fantasia 11 por el séptimo tono
d -2 -3 -4	Daza	Da011	Da 11.3						147	Fantasia 11 por el séptimo tono
d 1 -5 -3 -2 1	Daza	Da011	Da 11.4						436	Fantasia 11 por el séptimo tono
d -3 -2 1	Daza	Da011	Da 11.5						40	Fantasia 11 por el séptimo tono
d 2 3 4 4 3	Daza	Da011	Da 11.6						938	Fantasia 11 por el séptimo tono
d 1 1 -5 1 2	Daza	Da011	Da 11.7						488	Fantasia 11 por el séptimo tono

Fig. 5. Vihuela Database, tabla Fantasia Themes. Esteban Daza, Fantasía 11.

lidad, la base de datos incluye 923 personas, pero este número probablemente será reducido mediante la eliminación de algunos individuos marginales.

Para cada persona se dan detalles de las fechas de nacimiento y defunción o los límites cronológicos de su actividad, junto con lo que se ha podido saber sobre su localización geográfica, su profesión, estatus social y esfera social. Utilizando menús con valores preestablecidos en cuanto posible, estas clasificaciones permiten la definición y manejo de material referente a determinados grupos sociales. Sería fácil, por ejemplo, generar una lista de clérigos vihuelistas viviendo en monasterios, de mujeres vihuelistas, de violeros en Aragón en el último cuarto del siglo XV, o aficionados de urbanas de mediados del siglo XVI en Andalucía. Las combinaciones son casi infinitas, pero extremadamente útiles para el estudio del lugar de la vihuela en la sociedad española renacentista.

Las conexiones relacionales entre las personas en la tabla *Persons*, así como las conexiones con las de *Documents* y *Bibliography* la convierten en una herramienta bastante poderosa. Permiten vínculos directos entre múltiples individuos en la misma tabla. Es fácil, por ejemplo, mostrar los músicos que tocaban juntos o los miembros de una misma familia, o los estirpes de violeros como la Mofarriz, el Albariel o el Bejerano, y sus enlaces matrimoniales. El siguiente ejemplo muestra el registro de Hernando Bejerano, violero en Madrid en la segunda mitad del siglo XVI, con el elenco de sus parientes músicos a la derecha y, por debajo, los documentos donde se conserva la información primaria (Fig. 6).

La tabla *Documents* incluye todo tipo de fuentes escritas contemporáneas con los hechos. La mayoría son documentos notariales, pero también manuscritos conservados en bibliotecas de archivos, y los libros impresos durante la época. Se incluyen, entonces, los libros de música para vihuela junto a otros libros que ofrecen testimonio original pertinente. La inclusión de estos libros en dos tablas de la base de datos –documentos y bibliografía– es la única duplicación que ha sido necesario en todo el proyecto.

Aparte de los datos bibliográficos habituales y resúmenes de sus contenidos, la innovación principal de la tabla *Bibliography* son sus campos relacionales que

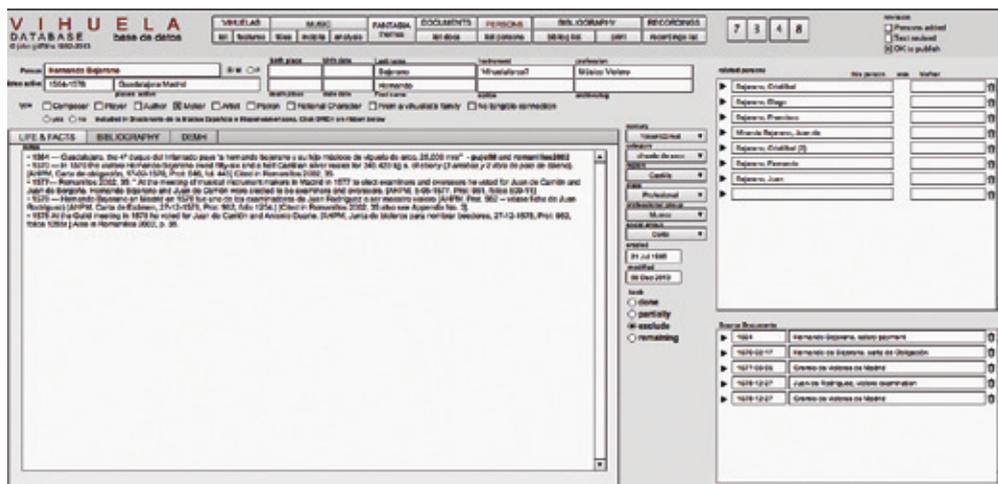


Fig. 6. Vihuela Database, tabla Persons. Registro de Hernando Bejerano.

producen referencias cruzadas a las tablas *Persons* y *Documents* para establecer la conexión entre los estudios musicológicos modernos, el material documental que manejan y las personas mencionadas. Es el caso de la edición moderna de Emilio Pujol (1949) de los *Tres libros de música* de Alonso Mudarra (1546) que muestra las referencias a personas y documentos (Fig. 7).

Por último, la base de datos discográfica, *Recordings*, registra las grabaciones de música de vihuela en instrumentos históricos y facilita la tarea de identificar y buscar lo que en la época moderna se ha grabado. Por otra parte, permite la reflexión sobre el desarrollo de los instrumentos y la práctica interpretativa en nuestro propio tiempo durante las primeras grabaciones de Emilio Pujol en 1935.

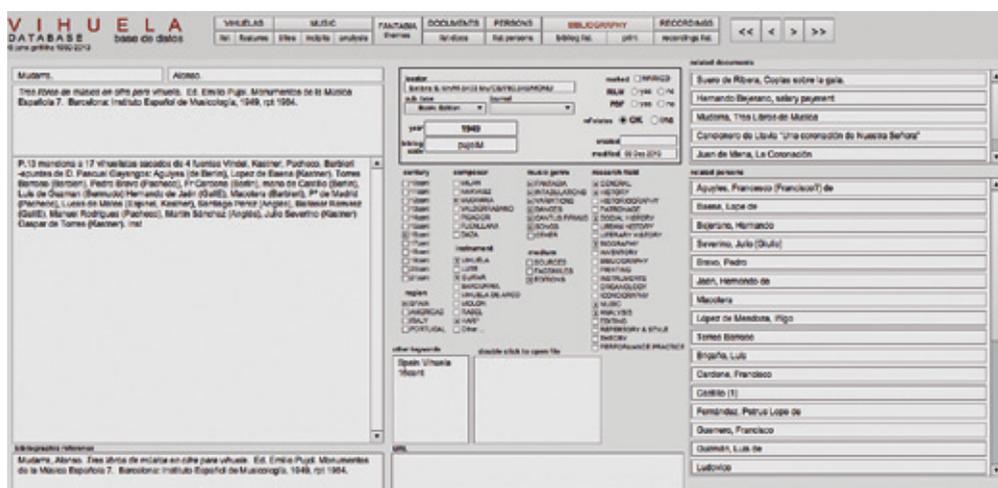


Fig. 7. Vihuela database, tabla Bibliography. Alonso Mudarra, Tres Libros de música.

En mi primer intento de rastrear esta historia, la base de datos me ha demostrado ser una herramienta valiosa (Griffiths, 2013). Los registros incluyen los detalles de cada grabación y artista, datos de su publicación, y una lista de las obras incluidas. Automáticamente, la información de cada grabación se enlaza a través de un vínculo relacional con la fichas correspondientes en la tabla *Music*.

BIBLIOGRAFÍA

- Bermudo, Juan, *Declaración de instrumentos musicales*, Ossuna, Juan de León, 1555, fol. 96.
- Griffiths, John, "Las vihuelas en la época de Isabel la Católica", *Cuadernos de Música Iberoamericana*, 20 (julio-diciembre 2010), pp. 7-36.
- Griffiths, John, "The Changing Sound of the Vihuela", en Kathleen Nelson y Maricarmen Gómez (eds.), *A Musicological Gift: Libro Homenaje for Jane Morlet Hardie*, Lions Bay (Canada), Institute of Mediaeval Music, 2013, pp. 123-146.
- Hyatt King, Alexander, "The Music Librarian, his tasks, national and international", *Fontes Artis Musicae*, 6 (1959), pp. 49-55.



RESUMEN

Este trabajo explora el uso de la informática aplicada a la musicología, en primer lugar, resumiendo una selección de proyectos representativos y claves en la área de las humanidades digitales internacionales, seguido por una introducción del proyecto que tengo sobre la mesa de crear una base de datos dedicado a la vihuela de mano, instrumento musical español renacentista por excelencia. La ponencia subraya el alto nivel de organización profesional en la musicología que permitió desde mediados del siglo XX la creación de proyectos internacionales de catalogación de fuentes, inicialmente en papel e ideales para su transformación en herramientas digitales. Explora algunos de los proyectos principales dedicados a la teoría musical, el canto llano, polifonía renacentista, y los nuevos desafíos informáticos para el manejo digital de información musical. La segunda parte se dedica a la creación de un hábitat digital para todo lo que se refiere a la vihuela: al instrumento y su música, a las personas, bien sean tañedores, constructores o consumidores, y la información documental, bibliográfica y discográfica relacionada.

Palabras clave: Música, canto llano, polifonía renacentista, motete, vihuela.

ABSTRACT

This article explores the use of computing applied to musicology, in the first instance, summarising a representative selection of key projects in relation to international digital humanities, followed by an introduction to my current project to create a database dedicated to the vihuela de mano, the Spanish renaissance musical instrument *par excellence*. The article underlines the high level of professional organisation in the discipline of musicology that permitted, from the middle of the twentieth century, the creation of international projects to catalogue the world's musical sources, firstly on paper but ideal for transformation into digital resources. The article explores reviews key projects related to music theory, plainchant, renaissance polyphony, and the new computational challenges for the digital handling of musical materials. The second part is devoted to the creation of a digital environment for the world of the vihuela: the instrument, its music, its players, makers and consumers, as well as documentation, bibliography, and related discography.

Keywords: Music, plainchant, Renaissance polyphony, motet, vihuela.

Centros y fronteras: el panorama internacional de las humanidades digitales

PAUL SPENCE

Department of Digital Humanities, King's College London
paul.spence@kcl.ac.uk, <<http://hd.paulspence.org/>>

Quiero agradecer a Carmen Isasi Pérez, Nàdia Revenga García y Antonio Rojas Castro por ofrecer comentarios sobre versiones anteriores de este capítulo.

INTRODUCCIÓN

A primera vista, el panorama de las humanidades digitales no parece nunca haber sido mejor a nivel internacional. Hay cada vez más publicaciones sobre el campo –que ha recibido bastante atención en la prensa a nivel internacional en los últimos años¹– y empieza a destacar en algunos países la formalización en estructuras académicas a través, por ejemplo, de cursos de posgrado o de cátedras². Pero los debates sobre su naturaleza y su alcance, lejos de apaciguarse, se han avivado en los últimos años a causa de los intentos de expansión de campo y la crisis financiera, con sus efectos conocidos sobre el ámbito científico. Se reavivan viejos debates sobre un campo que se transforma con relativa rapidez, y donde los centros y límites a niveles disciplinar e internacional de su actuación científica están bajo el microscopio una vez más. Este artículo pretende señalar algunos de los retos que le esperan todavía –en medio de transformaciones académicas y sociales significantes causadas por la era digital– y explorar cuáles pueden ser los modelos académicos adecuados para la investigación digital en humanidades.

En un escrito de septiembre de 2012, Luis Rodríguez-Yunta nos preguntó si el término ‘humanidades digitales’ es una mera etiqueta o un campo por el que deben apostar las ciencias de la documentación. Para los neófitos, el término ‘humanidades digitales’ ofrece tanto incógnitas como oportunidades para “pro-

¹ Un ejemplo muy conocido es la serie de artículos en el *New York Times*, que lleva la etiqueta ‘Humanities 2.0’ <http://topics.nytimes.com/top/features/books/series/humanities_20/index.html> [22/11/2013].

² Se han recibido noticias recientes sobre cátedras en humanidades digitales en las universidades de Leipzig, Aberystwyth, Passau, West Sydney.

yectar sus propios deseos... [y] ... sus fantasmas científicos" (Le Deuff, 2012 *–la traducción es mía*) y su difícil definición ofrece un amplio espacio para visiones heterogéneas sobre la relación compleja y difícil entre la tecnología y las humanidades. Una primera respuesta a la pregunta de Yunta, entonces, pasa por recordar que el campo, bien con el nombre actual o la denominación anterior, 'informática humanística', tiene una historia, unos cimientos que datan de mucho tiempo antes de la ola actual que tanto revuelo está causando en el mundo científico. Pero aquí nos encontramos con una ironía: con contadas excepciones, un campo que tanto ha contribuido a publicar, archivar y difundir la memoria cultural, no ha sabido narrar bien su propio pasado.

CENTROS Y FRONTERAS

Se suele atribuir el nacimiento de la 'informática humanística' al, encuentro, en 1949, entre el padre Busa y un fundador de IBM, que dio lugar a un proyecto de investigación de muchas décadas y a la obra *Index Thomisticus*, la labor de lematización de las obras de San Tomás de Aquino. Faltan estudios historiográficos sobre el período desde esta iniciativa de Busa hasta el día de hoy, que tendrían que contemplar varios ciclos en la historia del campo³:

1. Ciclos determinados por la tecnología y las capacidades funcionales que despierta: al principio, el acceso limitado a enormes ordenadores centrales para unos pocos privilegiados (contar datos, experimentación inicial, todo relacionado con aspectos de la computación lingüística); el nacimiento del ordenador 'personal', que abre el acceso a más personas (operaciones más variadas, incluyendo la preparación de textos y su publicación); el World Wide Web y el correo electrónico (que popularizan conceptos de la comunicación por red y la publicación digital rápida y abierta); Web 2.0 y los medios sociales (dos hechos que permiten compartir gustos e ideas, un paso más en la supuesta horizontalización de la comunicación, la importancia creciente del lector/usuario); y varios fenómenos posteriores, como la minería de datos, los datos grandes (*Big data*) y la web semántica que ofrecen nuevas maneras de captar, organizar y compaginar información estructurada de manera semántica a gran escala.
2. Ciclos humanos, que revelan la actitud social hacia estos cambios y su impacto en la investigación científica en general.
3. Y ciclos de investigación en humanidades digitales, de examinar cómo los investigadores han sabido adaptar la tecnología a las particularidades, y los retos, innovaciones y fracasos que esto ha provocado.

Sin querer predeterminar los resultados de tales estudios, podremos pronosticar que se notará el peso (a veces demasiado tradicional) de la filología y de las 'viejas' humanidades (por otro lado, un término nada despectivo) en su historia, aunque ahora esto está cambiando, y observamos la estrecha relación entre el flujo de las novedades tecnológicas y la evolución en la agenda científica de las humanidades digitales. Esta transformación continua en los puntos de enfoque y las fronteras epistemológicas del campo ofrece innovación (muy

³ El proyecto 'Hidden histories' es uno de los primeros intentos de trazar una historia del campo (Nyhan *et al.*, 2012).

atractiva para las ciencias académicas) pero también inestabilidad, lo cual dificulta una definición estable de la disciplina.

De la informática humanística a las humanidades digitales

El interés actual en las humanidades digitales, de mayor grado a la atención recibida hace unos años, se debe sin duda en parte a la mayor penetración general de la tecnología y la *cultura* digital en la sociedad, pero es probable que el cambio de nombre de 'informática humanística' a 'humanidades digitales' en varias lenguas y regiones del mundo⁴ haya jugado un papel determinante al abrir el abanico científico a más áreas de significado en el espacio entre la cultura humana y la tecnología. La palabra 'digital' ofrece una imagen menos *técnica* que el término 'informática', que une un formato de almacenamiento y transmisión binario de datos con algo humano, los dedos, poniendo algo muy técnico en manos de la sociedad y a la vez (al menos en castellano) parece poner más énfasis en 'las humanidades' al elevarlas de adjetivo a sustantivo. Este cambio de nombre, procedente de una decisión de marketing surgida a raíz de la publicación del compendio que sirve como hito refundacional de la disciplina (Kirschenbaum, 2010) –el Blackwell Companion to Digital Humanities (eds. Schreibman *et al.*, 2004), no solamente ha supuesto un mayor reconocimiento –imaginado o real–, sino que nos acerca más a otros campos centrados en la tecnología, como la cultura digital y las ciencias de la comunicación⁵.

Es llamativo que entre las entidades que actualmente forman parte de la alianza histórica de asociaciones en informática humanística/humanidades digitales, Alliance of Digital Humanities Organizations (ADHO), la gran mayoría usan el segundo término, una tendencia que continúa si incluimos a otras asociaciones más recientes/no inscritas en ADHO⁶, aunque esto no incluye a organizaciones como Humanities, Arts, Science and Technology Alliance and Collaboratory (HASTAC)⁷ que acaparan conceptos más generales de ciencia digital⁸, y aparte,

⁴ El desplazamiento semántico que ofrezco aquí es más matizado en algunos contextos, como sugiere Galina en su análisis de términos empleados a nivel histórico, que incluyen "Recursos Digitales para las Humanidades, Cómputo para las Humanidades, Cómputo en las Humanidades, Informática Digital y Cultural e Informática para las Humanidades" (Galina, 2011).

⁵ Estos campos "también incorporan puntos de encuentro con la comunicación, a través de las redes sociales o *community manager*, y las Ciencias Sociales, como la Biblioteconomía y Documentación." Baraibar, entrevista en *Diario de Noticias* (Navarra), 24 de mayo de 2013, 'Las Humanidades deben incorporar nuevas tecnologías, un investigador que no está en Internet casi no existe'.

⁶ Dentro de Alliance of Digital Humanities Organizations (ADHO), están The European Association for Digital Humanities (EADH, antes Association for Literary and Linguistic Computing – ALLC) <<http://eadh.eu/>>, centerNet (una red internacional de centros de humanidades digitales) <<http://digitalhumanities.org/centernet/>>, Australasian Association for Digital Humanities (aaDH) <<http://aa-dh.org/>>, Japanese Association for Digital Humanities (JADH) <<http://www.jadh.org/>>. En estos momentos, las siguientes asociaciones no están inscritas oficialmente, aunque algunas están en proceso de diálogo con ADHO: Humanidades Digitales Hispánicas (HDH) <<http://www.humanidadesdigitales.com/>>, Digital Humanities im deutschsprachigen Raum (DHD) <www.dig-hum.de/>, Red de Humanidades Digitales de México (RedHD) <<http://humanidadesdigitales.net/>>, Associação das Humanidades Digitais (AHDig) <<http://ahdig.org/>> [22/11/2013].

⁷ <<http://www.hastac.org/about>> [22/11/2013].

⁸ No obstante, es llamativo que el enfoque de HASTAC scholars a veces se alinea estrechamente con una agenda de humanidades digitales: <<http://www.hastac.org/blogs/herr-cafe/2013/10/30/what-hastac-scholars-2014-are-talking-about>> [22/11/2013].

vemos algunas excepciones, notablemente la asociación italiana AIUCD, que combina el nombre histórico, 'informatica umanistica', con 'la cultura digitale' en italiano⁹.

No parece, entonces, haber mucha duda sobre la tendencia hacia el uso de las 'humanidades digitales' como término, y la desaparición paulatina de 'informática humanística' y sus equivalentes en otras lenguas, pero ¿el campo en sí representa un fenómeno más o menos homogéneo a nivel internacional o una acumulación de prácticas y perspectivas relacionadas, y dónde están las grietas epistemológicas? ¿Siguen divisiones geográficas, lingüísticas o disciplinares, por ejemplo? Esto es el tema de la siguiente sección.

Hacia una definición de las humanidades digitales

¿Cómo podemos definir un fenómeno que, como todos acordaremos, ha gozado y sufrido de tantas transformaciones en su breve historia? Aquí analizo algunos de los puntos de encuentro, las formas de organizarse que pueden señalar los símbolos, las claves discursivas y las relaciones de afinidad de esta 'comunidad imaginada', apropiándonos de la expresión de Benedict Anderson (1991), y presento algunos de los criterios por los que podemos apuntar hacia una definición de las humanidades digitales.

Un aspecto notable en cualquier estudio sobre la terminología propia de las humanidades digitales es el uso extendido de la palabra 'comunidad', lo cual marca una diferencia con el lenguaje utilizado en otras disciplinas¹⁰. En parte se podría atribuir esto a una influencia de los conceptos de comunidades científicas o comunidades digitales de investigación ligados a la teoría de 'la red', pero también refleja, como varios han observado, un paradigma que depende mucho de voluntarismo, de una actitud de compartir (código, datos, publicaciones, a menudo con un compromiso fuerte con el Acceso Abierto), que a veces incluso tiene matices casi subversivos, de querer alterar o incluso derrocar algunas estructuras científicas actuales, o al menos de crear nuevas estructuras más apropiadas para una era humana dominada por los medios digitales. Y este concepto de comunidad está fuertemente relacionado con una realidad básica de la investigación en humanidades digitales; es decir, el hecho de que, en la mayoría de los casos, la investigación dependa de la colaboración, un hecho frecuentemente citado como un bien evidente e incuestionable, aunque recientemente estén surgiendo voces disonantes al respecto (McCarty, en prensa).

Una manifestación ejemplar de este fenómeno son 'las comunidades de práctica'. Tal es el caso de la comunidad para el estudio del siglo XIX, NINES¹¹, que pretende compartir ideas, herramientas, estándares para integrar la ciencia tradicional en la era digital, y, en algunos casos (como pasa con NINES), van

⁹ AIUCD significa, entonces, "Associazione per l'informatica umanistica e le cultura digitale". También salen de la tendencia ACH (Association for Computers and the Humanities) <<http://ach.org/>> y la rama francesa de la asociación canadiense CSDH-SCHN (Canadian Society for Digital Humanities/Société canadienne des humanités numériques [CSDH/SCHN]) <<http://csdh-schn.org/>> [22/11/2013].

¹⁰ Esto puede sorprender a colegas en otras disciplinas de las humanidades, más habituados al trabajo individual, o que no reconocen los supuestos detrás de la expresión.

¹¹ Networked Infrastructure for Nineteenth-Century Electronic Scholarship <<http://www.nines.org/>> [22/11/2013].

más allá y ofrecen servicios especializados, como, por ejemplo, la revisión por pares de investigación digital.

A veces estas comunidades tienen una estructura más formal, más fácil de identificar; o bien se trata de comunidades más informales –como puede ser la comunidad epigráfica digital que rodea el estándar de marcación digital EPIDOC¹² y la lista Digital Classicist¹³ (y que tiene lazos estrechos con investigadores en papirología)–. En ocasiones su relación con la ciencia digital está todavía en fase de negociación, como es el caso de CHARTA, un “proyecto destinado a la publicación en red de un corpus de textos y documentos antiguos de los siglos XII al XIX de España e Hispanoamérica”¹⁴. Pueden tener un enfoque disciplinar (los estudios decimonónicos como NINES o los estudios medievales, como Digital Medievalist¹⁵) o un enfoque sobre retos tecno-humanísticos generales como la edición de textos estructurados (TEI¹⁶) o los datos enlazados (ADHO SIG LOD¹⁷). Es evidente que estas comunidades virtuales representan un fenómeno global en la actividad científica y en la sociedad en general, pero en el caso de las humanidades digitales tienen un carácter sensiblemente práctico, que suele implicar a todo el ciclo de la investigación.

Varias asociaciones profesionales internacionales, algunas de ellas históricas¹⁸, han hecho esfuerzos mayores por promover estructuras semiformales en los últimos años, muchas dentro de la alianza ADHO (Alliance of Digital Humanities Organizations), sin perder el espíritu de colaboración que representa un principio fundamental del campo. Estas iniciativas, aunque sufren de un sesgo lingüístico y cultural hacia algunos países anglófonos que analizaremos más adelante, representan intentos de cubrir varias bases de la actividad científica, tanto *formales* –la publicación¹⁹, la organización de congresos²⁰, el reconocimiento científico²¹, el apoyo a jóvenes investigadores²², apoyo financiero a proyectos y talleres (EADH y ACH²³)– como *informales* –foros de discusión (Humanist²⁴), foros de consulta (Digital Humanities Questions & Answers)²⁵, y servicios de tutoría/apoyo laboral²⁶.

Estas asociaciones han jugado un papel decisivo en darle visibilidad a las humanidades digitales, uniendo aspectos formales –se puede interpretar como una respuesta concertada a la inquietud histórica sobre el reconocimiento académico oficial, que pretende crear todas las funciones oficiales de un campo– con espec-

¹² <<http://sourceforge.net/p/epidoc/wiki/Home/>> [22/11/2013].

¹³ <<http://www.digitalclassicist.org/>> [22/11/2013].

¹⁴ <<http://www.charta.es/>> [22/11/2013].

¹⁵ <<http://www.digitalmedievalist.org/>> [22/11/2013].

¹⁶ <<http://www.tei-c.org/index.xml>> [22/11/2013].

¹⁷ <<http://adho.org/announcements/2013/update-proposed-adho-sigs>> [22/11/2013].

¹⁸ Se fundaron la Association for Literary and Linguistic Computing (ALLC) –ahora European Association for Digital Humanities (EADH)– en 1973 y Association for Computers and the Humanities (ACH) en 1978.

¹⁹ <<http://adho.org/publications>> [22/11/2013].

²⁰ <<http://adho.org/conference>> [22/11/2013].

²¹ <<http://adho.org/awards>> [22/11/2013].

²² <<http://adho.org/awards/conference-bursary-awards>> [22/11/2013].

²³ <<http://ach.org/activities/grants-and-awards/>> y <<http://www.allc.org/research>> [22/11/2013].

²⁴ <<http://dhumanist.org/>> [22/11/2013].

²⁵ <<http://digitalhumanities.org/answers/>> [22/11/2013].

²⁶ <<http://ach.org/activities/mentoring/>> [22/11/2013].

tos innovadores, que se aprovecha de todas las posibilidades del medio digital dentro de un marco con fuerte tono 'comunitario'. No es de sorprender que, al menos en el mundo anglófono, haya habido bastante solapamiento de personal (y sinergia) con algunas comunidades de práctica como TEI, Digital Medievalist y NINES.

Mientras que las asociaciones ligadas a ADHO (y otras asociaciones internacionales no inscritas en ADHO) han sido creadas como entidades con estructura formal (aunque luego tienen expresiones informales e incluso descentralizadas como hemos visto), otras entidades (HASTAC, cuyo acrónimo se pronuncia como 'haystack', la traducción de 'pajar' en inglés, que enfatiza el aspecto informal de la comunidad) e iniciativas (como THATCamp, literalmente una desconferencia) priman una visión descentralizada, espontánea, que rehúye de los símbolos y tradiciones de los centros de poder universitario. Es dudoso que haya tanta diferencia entre estos dos modelos en la práctica –de nuevo se nota cierta repetición de nombres en los dos entornos– pero el enfoque social de esta segunda categoría, unido con su cobertura científica más general ("HASTAC is an alliance of more than 11,500 humanists, artists, social scientists, scientists and technologists working together to transform the future of learning for the 21st century"²⁷) nos llama a una reflexión más general sobre el impacto de la tecnología en la ciencia.

Es notable que esta segunda categoría suela promover y nutrirse de las humanidades digitales pero tiene menos tendencia a representarlas como un campo bien definido (nótese, por ejemplo, la palabra "transdisciplina" empleada en el 'Manifiesto por unas humanidades digitales' que salió del THATCamp de París en 2011²⁸, fuertemente influido por un actor importante en la pugna por una ciencia abierta, digital y conectada, es decir, el portal OpenEdition²⁹).

Si paso ahora a una reflexión sobre esas 'humanidades' moduladas (y torcidas) por ese adjetivo 'digitales', debo sugerir que sería interesante hacer un estudio a fondo de cómo diferentes definiciones y agrupaciones estructurales tienen influencia sobre el concepto de 'humanidades digitales' a nivel institucional o geográfico, pero por ahora podemos reconocer tendencias históricas hacia ciertas disciplinas: en Estados Unidos parece claro que ha primado la estrecha relación entre las humanidades digitales y los departamentos de Inglés (Kirschenbaum, 2010), mientras que en Europa algunos han observado una tendencia más bien hacia la historia y los estudios clásicos (Dacos, 2013). En general sigue habiendo una fuerte tendencia hacia la edición y las disciplinas o tecnologías textuales, tanto en el plano humanístico como en el técnico, y una división (a veces conflictiva) entre las humanidades 'tradicionales' y las ciencias más destinadas al estudio del mundo contemporáneo, como las ciencias de la información o las ciencias humanas con una perspectiva más social (la agrupación en 'ciencias humanas y sociales' parece favorecer esta tendencia), que apela a una reexaminación del papel de las humanidades, que trataremos más adelante.

Algunos ven en esto una contradicción entre las autodefiniciones del campo, reflejadas en las convocatorias de congresos o las declaración de la misión de las asociaciones –que ofrecen una visión abierta e inclusiva (la 'Big Tent' o 'carpa

²⁷ <<http://www.hastac.org/about>> [22/11/2013].

²⁸ <<http://tcp.hypotheses.org/487>> [22/11/2013].

²⁹ <<http://www.openedition.org/>> [22/11/2013].

Coruña en julio de 2013, y el congreso DH en Nebraska en el mismo mes. En los dos casos, predominan palabras generales de investigación o categorización de la información, como 'datos', 'investigación', 'archivo', 'proyecto' sobre palabras que reflejan el carácter crítico de las humanidades, o las propias disciplinas humanísticas (aunque hay algunas excepciones como 'la historia').

Si es difícil llegar a una definición estable del campo que nos ocupa, tal vez sea más fácil acordar primero lo que no es. Si las humanidades digitales pretenden ser algo más que 'las humanidades en la era digital', es decir las humanidades estudiadas y enseñadas con herramientas digitales y en un entorno al menos parcialmente digital, no deben incluir *automáticamente*, a mi juicio, el mero hecho de digitalizar un material de estudio, o el mero uso de herramientas digitales, como escribir un blog, publicar una página web o crear una base de datos (aunque estas actividades pueden formar perfectamente parte de la investigación en el campo cuando realmente aportan algo nuevo a nivel científico). Esto no lo digo con ningún afán de excluir, ni tampoco de elitismo digital (de hecho hay que evitar a toda costa el complejo de superioridad representado en la imagen muy difundida que reza 'Yo hago humanidades digitales. Es como lo que tú haces, pero mejor'³¹), sino de proteger y mantener el sentido histórico de las humanidades digitales: de ser innovadores y críticos en el uso de la tecnología digital para las humanidades que a menudo rehúye del concepto de 'productos' digitales estables. Aunque depende frecuentemente de procesos de digitalización (o los tiene como resultado suplementario) no se trata simplemente de mejorar el acceso o la diseminación, sino de estudiar los (y a veces crear nuevos) modos de crear y gestionar el conocimiento. En un ensayo sobre el 'desasosiego productivo' de la ciencia digital en el siglo XXI, Julia Flanders (2009) nos recuerda que no se trata del descubrimiento o de la 'solidez' de lo conocido, sino de representar (con modelos digitales) ese conocimiento y en algunos casos de representar el proceso de investigación: Flanders añade que es un examen de cómo sabemos cosas, y de cómo las presentamos para el estudio.

EL FUTURO DE LAS HUMANIDADES DIGITALES: RETOS Y OBJETIVOS

En la primera parte de este ensayo he analizado algunos elementos que han contribuido a definir las humanidades digitales hasta ahora, pero ¿cuáles son los retos principales que ha de afrontar en el futuro?

Proposición 1: definirse mejor ante otras áreas, disciplinas, comunidades próximas

La dificultad para llegar a definiciones consensuadas del campo a causa de su elasticidad, como hemos notado antes, ha sido complicada, sin duda, por los propios intentos de ampliar sus parámetros con un nuevo nombre y los retos (y oportunidades) presentados por el interés reciente de amplios sectores de la ciencia. Esto a veces ha creado tensiones, y conforme estos límites epistemológicos se mueven o se expanden, se percibe un cambio significativo en los

³¹ Entrada de blog enero de Jean-Christophe Plantin (2012), <<http://cartonomics.org/2013/01/12/la-these-dure-trois-ans-amours-et-desamours-de-la-these-contemporaine/>> [22/11/2013].

centros discursivos de las humanidades digitales. Aunque los debates suscitados a partir de varias ponencias en los últimos años alrededor del Modern Language Association of America (MLA) tienen un tono fuertemente norteamericano³², las inquietudes expresadas por Stephen Ramsay (2013) en sus entradas de blog 'DH Types One and Two' son simbólicas de una división interpretativa importante entre un primer grupo que se siente unificado por una serie de metodologías (que incluye la marcación digital, el análisis textual, la visualización, el análisis geoespacial, la minería de datos y el análisis semántico) empleadas para mejorar el estudio de la memoria cultural, y un segundo grupo más amplio que, según la definición de Ramsay, incluye a investigadores en cultura digital y las ciencias de comunicación o documentación, críticos culturales con interés en el internet y artistas digitales. En su interpretación, el primer grupo representa una comunidad reconocible, mientras que el segundo grupo es más difuso, y a veces refleja "cierta disposición revolucionaria" que se habría apoderado del universo científico.

Otra manera de trazar las diferencias es que la primera tendencia pretende, de una manera bastante práctica, reconfigurar la investigación en humanidades a través de entornos y herramientas digitales, mientras que la segunda pretende convertir las humanidades en un debate sobre el estado humano en la era digital. No existen necesariamente líneas divisorias claras entre las dos concepciones en todos los casos, pero las tensiones intelectuales ya existentes son testimonio de la necesidad de buscar puntos de conexión y establecer límites, donde sea necesario, para no perder la definición.

Los focos disciplinares de estos diálogos e intercambios epistemológicos son difusos, pero a cierto nivel representan un éxito para las humanidades digitales que les servirán en el futuro para responder a críticas históricas (y en buena medida justas) vertidas hacia el campo sobre su falta de perspectiva crítica ("Where is criticism in the digital humanities?" Gibbs, 2011³³) y para reparar grietas en su relación práctica y discursiva con otras áreas científicas contiguas, como las bibliotecas digitales (¿por qué hay tan poca investigación sobre la relación científica entre bibliotecas digitales y humanidades digitales?) y sobre la investigación en cultura digital. Volviendo sobre este último punto, es chocante la falta de contacto histórico entre las humanidades digitales y los campos que estudian el impacto de la cultura digital, a pesar de las advertencias de figuras como Rockwell ("no debemos dejar la teoría digital a los investigadores de los Nuevos medios de comunicación"³⁴) y de Svensson ("una falta relativa de compromiso con lo 'digital' como objeto de estudio" *-la traducción es mía*, 2009), y aunque eso está cambiando, en parte porque destacados investigadores en la teoría cultural como Pierre Levy han identificado a las humanidades digitales como un campo de interés³⁵, todavía les falta a las humanidades digitales acomodarse ante una rica historia científica que nos tiene mucho que enseñar.

³² Criticados por Andrew Prescott en su entrada de blog 'Small worlds and big tents' (2013).

³³ Gibbs responde, en parte, a un escrito de Alan Liu (2011) que identifica la falta de perspectiva crítica cultural.

³⁴ "Digital theory should not be left to new media scholars, nor should we expect to get it right so that we can go back to encoding of other humanities disciplines" (*la traducción es mía*). Rockwell (2004) citado en Svensson (2009).

³⁵ <<http://www.scoop.it/t/digital-humanities-for-beginners>> [22/11/2013].

A la vez, y ahora en lo que se refiere a lo estructural, se percibe la necesidad de agilizar los puntos de contacto y comunicación entre las formas clásicas de organización dentro de los departamentos tradicionales (o los centros o departamentos de humanidades digitales donde existen) y los MediaLabs, laboratorios y otras iniciativas que no siempre se identifican explícitamente con las humanidades digitales, y, aunque lo hagan, siguen una agenda más general de ciencia e innovación digital³⁶.

Estos debates, choques e intercambios definirán el futuro de las humanidades digitales, que a su vez tendrán que canalizar efectivamente su discurso, negociando vertientes muy dispares si no quieren ser asimiladas por otras disciplinas. El discurso que falta debe distinguir el papel de las humanidades digitales de debates más generales sobre la ciencia en la era digital y responder a la pregunta: ¿dónde están los puntos de diferencia, y por qué requieren una actividad más allá de lo puramente técnico? Sin duda la existencia de un concepto como las humanidades digitales (es llamativo que no se oiga hablar de ciencias sociales digitales o ciencias naturales digitales, por ejemplo) se debe no solamente a cierta lentitud en las humanidades a adaptarse a la era digital, sino también a la dificultad (a veces imposibilidad) de adaptar las metodologías o herramientas digitales provenientes del mundo comercial (o de otras ciencias) a una ciencia marcada por la interpretación (que se escapa de definiciones estables o cálculos cuantitativos) y el enfoque sobre detalles pequeños, y que por consiguiente requiere de metodologías y herramientas creadas por y para los humanistas.

Nos podríamos preguntar si es útil para una definición de las humanidades digitales la definición de Laura Borràs (2011) del internet como una 'interzona' que "simultáneamente sirve y trastorna las relaciones humanas" y que "altera nuestros procesos creativos y genera disciplinas nuevas que tendrán que coexistir con las disciplinas ya existentes" (la *traducción es mía*). Antes de llegar a un término final o al menos estable, este trastorno en las relaciones convencionales, esta generación de nuevas estructuras científicas, puede pasar por lo que Peter Galison denominó *Trading Zones*, procesos de canje entre dos culturas científicas que tienen que crear lenguajes intermedios y temporales, "pidgin" o "criollos" para comunicarse sin perder su propia identidad (Galison, 1997).

Este término ha sido introducido en discusiones sobre las humanidades digitales por Willard McCarty (1999) y después seguido por otros como Patrik Svensson (2011, 2012) para explicar su interacción con otras áreas científicas en un modelo que no necesariamente requiere una inmersión total de una comunidad científica en el modelo de otra, y donde el lenguaje es un elemento, una herramienta clave en estas zonas de intercambio. Aunque este modelo de zonas de intercambio es muy atractivo para un campo tan interdisciplinar como las humanidades digitales, un reto clave para el campo es cómo mantener un discurso que sabe combinar (y contrastar) los lenguajes humanísticos y lenguajes técnicos de una manera que aporta innovaciones a dos culturas académicas muy distintas. Se trata sobre todo de avanzar y afinar la agenda científica del campo, el tema del siguiente apartado.

³⁶ Pienso en grupos e iniciativas como Open Edition (<<http://www.openedition.org/>>), GrinUGR (<<http://grinugr.org/>>), Medialab USAL (<<http://medialab.usal.es/>>) y CCCB LAB (<<http://blogs.cccb.org/lab/es/el-lab/que-es/>> [22/11/2013]), entre otros.

Proposición 2: elaborar, fortificar una agenda científica independiente

Las humanidades digitales han contribuido en varios modos a una infraestructura de ciencia digital en humanidades, han sido utilizadas a menudo como anexo o soporte a la investigación en otros campos de las humanidades, pero con frecuencia su aportación independiente a la investigación ha quedado oculta o marginada. En parte esto se debe a estructuras institucionales que reducen el papel del campo a un 'servicio técnico', es decir a un 'apoyo a' en vez de 'una colaboración en' la investigación, y las dinámicas de los proyectos financiados que dejan poco tiempo o recursos para la investigación propia no ayudan, pero también hay que reconocer que los humanistas digitales han sido lentos en difundir su investigación en muchos casos, dependiendo mucho de comunicaciones de carácter más informal (los blogs por ejemplo). Hay pocos monográficos en humanidades digitales todavía en cualquier idioma, y aunque es válido cuestionar el predominio de los formatos tradicionales en la ciencia, al menos hay que ofrecer alternativas creíbles de modos de publicación más allá de los recursos creados como fruto de la investigación. Y finalmente esta agenda científica debe, en palabras de Willard McCarty en su presentación para el premio Busa en el congreso DH2013 (en prensa), tener "resonancia con las culturas intelectuales de las artes y humanidades", es decir ser reconocible como una aportación clara, colindante y creíble al estudio de la condición humana. En lo humanístico debe nutrir y nutrirse de las humanidades tradicionales, y en lo técnico debe superar el nivel de una mera implementación: el papel de las humanidades digitales es el de sacar conclusiones de cómo la tensión entre las humanidades y la tecnología transforma nuestra investigación, para bien y para mal (volviendo a McCarty, "no debemos salir ilesos del contacto con la tecnología... si la tecnología no nos cambia, nos encarcela").

Un observador externo podría preguntar por qué, sesenta años después de sus orígenes, el tema de la agenda no está más claro. En parte lo podemos atribuir al largo proceso de gestación de los proyectos digitales (donde corremos el peligro de invertir demasiado tiempo y emoción en nuestras invenciones, y se puede alegar, con razón, que los recursos digitales creados a menudo tienen equivalencia intelectual en su composición y razonamiento a un artículo o un monográfico). Pero a la vez podemos observar que el aspecto colaborativo, tan importante en la germinación de la investigación innovadora, dificulta la generación de ideas identificables como aportaciones individuales a la ciencia, y que el concepto de comunidad, de a veces casi estar 'entre amigos', no siempre conduce al pensamiento crítico. "La investigación no depende de la armonía, sino de debates controvertidos" afirma Thaller en su introducción a una edición especial dedicada a 'las controversias en las humanidades digitales' (2012)³⁷.

A veces se nota también cierta falta de perspectiva histórica. En las muchas ocasiones en las que he servido de evaluador de propuestas para congresos en humanidades digitales, el fenómeno más común en las propuestas de ponencia, sin importar en qué país se lleve a cabo el evento, es el temido 'informe de proyecto', en el que apenas se hace reflexión crítica sobre el estado de la cuestión hasta el momento, y en el que no se muestran indicios de conocer

³⁷ "Research is not driven by harmony, but by controversial discussion" (ed. Thaller, 2012: 9 -*la traducción es mía*).

bien el trabajo de otros en la misma área, o bien contiene bibliografía escasa o autoreflexiva. Hay que saber *construir* recursos, pero también hay que saber *deconstruirlos*.

La creación de una infraestructura digital ha sido un elemento importante en la historia de las humanidades digitales³⁸, pero esta no es en sí una actividad crítica o creativa. Un elemento importante del humanista digital, a mi juicio, es evitar la falsa brecha entre practicantes digitales y no digitales (en realidad todos operamos en entornos digitales y no digitales), así como evitar no solamente *los jardines amurallados* de productos digitales principalmente destinados a fomentar el consumo, sino también visiones tecno-positivistas que pretenden reducir la expresión cultural a un 'problema de especificación técnica'.

Y por último, quizás el reto más difícil para los humanistas digitales es cómo resolver la disparidad entre los modelos digitales y la perspectiva interpretativa de las humanidades tradicionales, donde el contexto es primordial. Un texto está ligado a su contexto cultural de lectura y de producción y no hay que separarlos, nos recuerda Bertrand (2012: 9), y hay que reconceptualizar los datos y los instrumentos creados más allá de etiquetas y campos en una base de datos, en un proceso que tendrá que contemplar una transmisión compleja y continua entre envases digitales y no digitales (Bertrand, 2012: 12). Se trata de compaginar la necesidad del ordenador de recibir y procesar información en segmentos bien definidos con la necesidad del ser humano, al menos con pensamiento crítico, de poder sugerir, intuir y ser ambiguo de manera constructiva. Navegar esta tensión, que puede ser una tensión productiva, es una de las tareas principales de las humanidades digitales.

Proposición 3: crear programas estables de formación y enseñanza

Históricamente las humanidades digitales han dependido de proyectos de investigación cuyo modelo de financiación ha tenido un efecto determinado (es decir restrictivo) sobre las estructuras laborales y la capacidad productiva de los investigadores. Es muy evidente, pero un campo que pretende aportar algo significativo en términos científicos necesita un programa independiente (o al menos identificable) de formación a todos los niveles, desde grado hasta doctorado. Ha habido varias experiencias a nivel internacional, con más éxito en programas de Máster que en grado o doctorado, pero es difícil sacar conclusiones de un panorama algo confuso. Se puede observar que ha sido más fácil crear programas de docencia en países que ofrecen mayor flexibilidad estructural, donde los programas han solido combinarse con otras disciplinas. Esto es algo natural en un campo de naturaleza interdisciplinar, y cuya naturaleza ha variado mucho desde cursos con un carácter muy práctico hasta cursos muy teóricos, pero en una situación donde, "los humanistas que no dominen las destrezas necesarias para tomar ventaja de las nuevas tecnologías digitales se arriesgan a ser marginados dentro del nuevo ecosistema digital que está surgiendo", como asevera Suárez (2013) las humanidades digitales deben desempeñar un papel significativo.

³⁸ Ver Borgman (2009) o el proyecto DARIAH <<http://dariah.eu/>> [22/11/2013].

Proposición 4: identificar los objetos académicos divulgativos y evaluables en las humanidades digitales, y crear los mecanismos apropiados de evaluación

Este apartado presenta tres tipos de pregunta:

1. ¿Cuál es un resultado científico válido para las humanidades digitales, digno de evaluación como equivalente a una publicación tradicional? ¿Basta crear (por muy sofisticadas que sean), una base de datos para representar una prosopografía tradicional, una edición digital en TEI/XML o una visualización mostrando relaciones literarias en la época de la ilustración? La dependencia mayor de las humanidades digitales de proyectos de investigación para abastecerse intelectual y financieramente ha contribuido a la visibilidad del campo, pero en cierto sentido también ha sido un lastre en términos de reconocimiento académico, y no solamente por las dificultades ya conocidas de los mecanismos de evaluación académica para adaptarse al paradigma digital.
2. ¿Cómo debemos, como humanistas digitales, presentar nuestra investigación para que sea reconocida por la comunidad científica, y para que llegue al público más amplio posible? Ya he sugerido que hay que hacer más esfuerzo en distinguir la aportación científica, y podemos añadir aquí que a menudo los proyectos creados han solido sacrificar el aspecto divulgativo por el intento de crear modelos digitales o interfaces impresionantes. Es cierto que hay resistencia en algunas áreas de las humanidades para reconocer nuevas formas de investigación, pero a la vez debemos reconocer que, frecuentemente, las innovaciones que podemos aportar han quedado encerradas en canales internos (o al menos informales) de comunicación, de difícil acceso para el neófito o el investigador de otros campos, y con frecuencia se estructura el hilo narrativo del argumento en términos puramente tecnológicos, en vez de centrarse en las preguntas básicas de la investigación humanística.
3. Más allá de cuestiones de definición de los objetos constituyentes de una agenda científica, ¿qué es lo que podemos evaluar, y cómo vamos a proceder a la evaluación en investigación interdisciplinar? Para tomar un ejemplo que conozco bien, en el proyecto *Out of the Wings*³⁹ hay una amplia variación de áreas de pericia, tanto a nivel humanístico (historiografía del teatro, traducción, representación teatral, edición) como a nivel técnico (modelos conceptuales de la información humanística en UML⁴⁰ y FRBR⁴¹, información estructurada en bases de datos, interfaz, participación del usuario, marcación digital en TEI/XML, por nombrar algunos elementos). En muchos lugares resuena la observación de Nagaski y Muller (2012) sobre que los mecanismos actuales de crédito entre las TIC y las humanidades no son compatibles, por lo que es difícil que un humanista consiga que el aspecto digital de su investigación sea evaluada,

³⁹ <<http://outofthewings.org/>> [22/11/2013].

⁴⁰ Unified Modeling Language, un lenguaje muy usado en la ingeniería de software.

⁴¹ Functional Requirements for Bibliographic Records, un modelo bibliográfico para captar información sobre la interacción con catálogos y bases de datos bibliográficas desde la perspectiva del usuario.

pero incluso si conseguimos llegar a un punto de reconocimiento mínimo en los sistemas formales de evaluación, ¿cómo se deben decidir los conocimientos adecuados para proyectos de investigación que contemplan varias disciplinas y varias habilidades?

4. Y finalmente ¿cuáles son los mecanismos apropiados para evaluar de forma justa el papel de los distintos participantes en un proceso de investigación de carácter colaborativo? El manifiesto de los jóvenes investigadores en humanidades digitales⁴² señala la brecha que se abre entre las prácticas digitales y su reconocimiento institucional, y ha habido varios intentos de afrontar el problema, entre ellos la iniciativa FairCite, que propone “una cultura más justa de citación en el ámbito académico” que contempla: aspectos de reconocimiento oficial a investigadores fuera de las estructuras formales de la investigación, estudiantes y personal contratado como ‘técnicos’; criterios para distinguir entre ‘autoría’ y ‘contribución’ en un proyecto de colaboración; así como normas/culturas para determinar el orden de una lista de nombres⁴³.

Proposición 5: formalizar su lugar en las instituciones académicas, respetando la autonomía de cada ‘foco organizativo’ y sus condiciones locales

Los debates sobre el estado ontológico de la informática humanística/las humanidades digitales y el lugar que les corresponde en las instituciones académicas llevan al menos quince años, y no parece haber más consenso que en el primer día en que se planteó esta cuestión. Es difícil, además, evitar la sensación de que algunos toman una postura concreta según sus intereses y su realidad institucional, y sin duda cuestiones nacionales e incluso locales tienen una influencia importante, pero ¿podemos sacar conclusiones generales?

Primero hemos de preguntarnos si las humanidades digitales siguen siendo el “comerciante fenicio entre naciones asentadas” que Willard McCarty describió en su análisis de la (entonces) informática humanística como interdisciplina en 1999. ¿No se debe vislumbrar un movimiento del campo desde la periferia al centro de las instituciones académicas, de un estado ambulante a un estado más fijo debido a la atención (incluso se puede decir el reconocimiento) pública e internacional recibida?

Ha habido varios intentos de ‘asentarse’ de manera formal en el mapa académico e institucional, y podemos observar varios estados o fases (ver el esquema abajo) en este proceso (con movimientos posibles en todas las direcciones). Es tentador trazar una evolución desde la actividad más informal (representada en la primera fila de la tabla) hasta el pleno reconocimiento representado por un departamento académico (ver última fila), pero en realidad se nos presentan varias gradaciones estructurales, que incluyen diferenciaciones entre el carácter ‘servicio’/‘académico’, el grado de reconocimiento oficial de la actividad a nivel institucional, el grado de reconocimiento académico por la comunidad científica,

⁴² *Young Researchers in Digital Humanities: A Manifesto* <<http://dhdhi.hypotheses.org/1855>> [22/11/2013].

⁴³ <<http://faircite.wordpress.com/>> [22/11/2013].

el balance docencia/investigación y los modelos financieros que implican, y el nivel de control sobre su propia agenda científica⁴⁴.

Tabla 1. Las humanidades digitales en las instituciones

<i>Estado institucional</i>	<i>Manifestaciones</i>	<i>Observaciones</i>
Estado latente	<ul style="list-style-type: none"> • Actividad individual. • Actividad interdisciplinar integrada en actividad de servicio/apoyo técnico, por ejemplo 'Computing Centre'. 	<ul style="list-style-type: none"> • Poca visibilidad. • Poco espacio formal para la investigación; depende del voluntarismo. • Visión 'tecnicista' de la actividad • Predomina el aspecto 'servicio'.
Actividad interdisciplinar	<ul style="list-style-type: none"> • Actividad interdisciplinar integrada en otra unidad/centro/departamento. 	<ul style="list-style-type: none"> • Más visibilidad, pero es aneja a la unidad 'padre'. • Este enfoque sobre los objetivos de la unidad que le cobija sirve muy bien a la agenda académica de esta, pero le falta agenda independiente. • Balance 'académico' o de 'servicio' por definir.
Foco interdisciplinar	<ul style="list-style-type: none"> • Centro de humanidades digitales (también Centro de informática humanística o Centro de 'actividad X' digital o Laboratorio). 	<ul style="list-style-type: none"> • Visibilidad completa. • Reconocimiento dentro de una dinámica interdisciplinar. • Relativa libertad para definir agenda de investigación. • Balance 'académico' o de 'servicio' por definir.
Departamento oficial	<ul style="list-style-type: none"> • Departamento de Humanidades Digitales. 	<ul style="list-style-type: none"> • Visibilidad y reconocimiento formal pleno. • Agenda académica independiente.

En Estados Unidos, en parte fomentada por la creación de una oficina de humanidades digitales en 2008⁴⁵, se ha visto la propagación de nuevos centros de humanidades digitales en la última década, pero en otros países se nos presenta un cuadro algo confuso que merece un estudio a fondo. Extendiendo la mirada hacia su contexto institucional un poco más, nos podemos preguntar ¿a qué área académica global debe pertenecer la actividad en humanidades digitales? ¿Qué efecto han tenido, o deberían tener, las transformaciones digitales

⁴⁴ Tal vez sea útil ahora observar que mi propio departamento, el "Department of Digital Humanities", de King's College London, ha pasado por varias de estas etapas, naciendo como un centro de informática que compartía funciones de servicio y de investigación, luego creando en 1989 un programa de enseñanza que incluía informática humanística, estableciéndose el Centre for Computing in the Humanities en 1992, para terminar convirtiéndose en departamento en 2001/2002 (aunque no adoptó su nombre actual, "Department of Digital Humanities" hasta 2011, el mismo año en que incorporó el antes independiente Centre for e-Research -CeRch-).

⁴⁵ National Endowment for the Humanities/Office of Digital Humanities <<http://www.neh.gov/divisions/odh>> [22/11/2013].

sobre la universidad, y qué implicaciones tiene esto para las humanidades digitales? En términos disciplinares ¿es mejor que la humanidades digitales formen parte de la facultad de letras, de filología, de informática, de biblioteconomía, de comunicación y documentación, de cultura digital, o a ninguna? Es evidente que el concepto local o nacional/regional de 'las humanidades' será un factor determinante a la hora de responder a esta pregunta.

Pasando de lo institucional a lo individual, un investigador en humanidades digitales puede tener muchos papeles, que van desde lo muy técnico (programación) hasta lo muy 'humanista', pero según mi experiencia hay poca relación entre el título oficial y su función real en el campo, sobre todo si se trata de su grado de aporte a la ciencia⁴⁶. La cultura de los títulos profesionales en humanidades es uno de los terrenos donde más se notan los criterios locales (ramas 'académic'/'researcher' en Reino Unido, ramas 'chercheur'/'ingénieur' en Francia etcétera), pero en el fondo las diferencias son parecidas en la mayoría de los países: una fuerte diferenciación entre las cabezas visibles académicas ('los investigadores principales' en el lenguaje de las becas de proyecto) y los otros investigadores, con, además, una nomenclatura que distingue entre 'humanistas' y 'técnicos' (que evoca un contraste entre 'ciencia' y 'servicio' en muchos contextos), todo muy poco representativo de la realidad actual en la investigación misma. Para que una colaboración maximice su potencial es importante que la interacción sea bidireccional (los criterios son humanísticos pero las pautas tecnológicas determinan lo que es posible llevar a cabo), iterativa (hacen falta revisiones de los dos lados para que la negociación entre *techne* e interpretación humana se resuelva satisfactoriamente) y la contribución, en términos de investigación académica, tiene que ser debidamente reconocida (Bradley, 2009)⁴⁷.

En los Estados Unidos ha tenido cierto eco el proyecto '#alt-ac' o '#alt-academy', que pretende trazar líneas de promoción laboral alternativas en el mundo académico⁴⁸, y los modelos financieros son un factor determinante –el 'dinero blando' del proyecto de término fijo, motor inicial de colaboraciones e innovaciones, viene a ser una cuestión problemática cuando se quiere formalizar un programa de investigación, mientras que la falta de creación de plazas fijas en las instituciones académicas acentúa las dificultades en mantener la continuidad intelectual de un programa de investigación. Todo esto apunta a la necesidad de promover nuevas líneas profesionales con una reflexión profunda y con vistas a largo plazo sobre los papeles típicos en colaboraciones en humanidades digitales (y su relación con los sistemas de crédito antes mencionados), los títulos profesionales y su relación con las estructuras en humanidades en general.

Proposición 6: crear unas humanidades digitales verdaderamente globales

Mi última reflexión sobre el futuro de las humanidades digitales a nivel global es precisamente su aspecto 'global', porque mientras que existe, sin duda,

⁴⁶ También es interesante notar que, sea cual sea su perfil actual, un porcentaje alto proviene de carreras humanísticas. Es el caso de DDH, por ejemplo, donde la mayoría de los investigadores en DDH a lo largo de su historia hicieron carreras en humanidades.

⁴⁷ Bradley dice "The scholar cannot expect to be the master of all the technologies required... our experience at CCH has shown us that the innovation in fact flows in both directions. The interaction has more the aspect of work between different but somewhat equal collaborators".

⁴⁸ <<http://mediacommons.futureofthebook.org/alt-ac/>> [22/11/2013].

una larga historia de humanidades digitales en países no-anglófonos, ni esta actividad ha recibido el mismo grado de reconocimiento en el campo a nivel internacional, ni tiene la misma coherencia estructural que tiene en algunos países anglófonos (con muy pocas excepciones). Aunque ha habido avances importantes en el tema de la representación cultural y lingüística, quedan todavía muchos escollos para una comunidad que pretende ser global e inclusiva. Se habla mucho, como hemos visto, de la comunidad de humanidades digitales en primera persona plural, pero “¿quién es ese ‘nosotros’?” pregunta Isabel Galina (2013). Esta comunidad se ha dado cuenta de que no es tan abierta o universal como se pensaba en términos de representación étnica, género, raza, lengua o clase, continúa, y hasta hace poco, en vez de preguntar ‘¿dónde está la crítica cultural en humanidades digitales?’ (Liu, 2011), podíamos preguntar ‘¿dónde está la crítica multicultural en las humanidades digitales?’

En su provocativa crítica cultural de las humanidades digitales en 2012, Domenico Fiormonte nos preguntó si existían ‘unas humanidades digitales no-anglo-americanas, y en caso afirmativo, cuáles eran sus características’⁴⁹. La rápida evolución de asociaciones regionales o nacionales en el panorama internacional de los últimos años (Galina, 2013b), junto con un fuerte enfoque multicultural en varios debates y foros (que incluyen los congresos internacionales de ADHO⁵⁰) apunta a una transformación en términos formales de la representatividad, pero no se deben pasar por alto unas cuestiones de fondo. Está claro que *hay trabajo por hacer*.

En realidad las asociaciones internacionales de humanidades digitales se han esforzado por afrontar una serie de cuestiones lingüísticas y culturales desde hace varios años a través de iniciativas como el comité de multilingüismo y multiculturalismo⁵¹ y proyectos como ‘digital diversity in practice: plotting language similarity’⁵², y ha habido otras iniciativas importantes como los congresos CLiP⁵³. Una vez superados los peligros de formulismo o populismo que estos debates pueden acarrear, nos enfrentan cuestiones lingüísticas y culturales complejas y enmarañadas. Para empezar, ¿cómo conseguir información objetiva sobre la representación, y qué representa precisamente?

Ha habido varios intentos de analizar e interpretar el estado del campo a nivel internacional⁵⁴, algunos utilizando datos ya existentes y otros creando nuevos sistemas para contabilizar datos de representaciones geo-culturales, que no nos ayudan todavía a formar una visión objetiva del campo, pero sí sirven para crear debate o identificar posibles áreas de investigación futura. Es el caso,

⁴⁹ “Is there a non Anglo-American Digital Humanities (DH), and if so, what are its characteristics?” (Fiormonte, 2012 *-la traducción es mía*).

⁵⁰ El tema para el congreso DH2014 (<<http://dh2014.org/>> [22/11/2013]) “Digital Cultural Empowerment theme” es una continuación natural del tema “Digital Diversity: Cultures, languages and methods” del congreso DH2012 conference en Hamburgo <<http://www.dh2012.uni-hamburg.de/>> [22/11/2013].

⁵¹ <<http://adho.org/administration/multi-lingualism-multi-culturalism>> [22/11/2013].

⁵² <<http://www.dh2012.uni-hamburg.de/conference/language-similarity-through-correspondence-analysis/>> [22/11/2013].

⁵³ <http://legacy.cch.kcl.ac.uk/clip2006/content/practicalities/past_conferences.html> [22/11/2013].

⁵⁴ Ver, por ejemplo, <<http://digihum.mcgill.ca/~sgs/dci-survey>>, o mis propios intentos de recopilar información sobre el campo en países castellanoparlantes <<http://hd.paulspence.org/recursos/hh-dd-es/>> [22/11/2013].

por ejemplo, de estudios como el hecho por Frédéric Clavert ('no ha ocurrido todavía la revolución multicultural en humanidades digitales'⁵⁵) que plantea temas importantes en el plano lingüístico y cultural (aunque a menudo refunden cuestiones de lugar de nacimiento, de origen, de residencia, lingüísticas y de afiliación institucional). Nos urgen estudios sobre el aspecto internacional de las humanidades digitales, tanto en el aspecto cuantitativo como en lo cualitativo, que traten a fondo lo geográfico, lo cultural y lo lingüístico.

En otro estudio sobre la constitución geográfica del campo, basado en una encuesta a 850 investigadores en línea ('¿Quiénes sois, humanistas digitales?'), Marin Dacos nos pregunta si el gobierno de humanidades digitales es equilibrado y democrático (2012). No cabe duda que la cara visible de las humanidades digitales muestra una dominación anglófona⁵⁶, y aunque existen casos (como EADH) donde la representación geográfica es más equilibrada, en general los mecanismos actuales de las entidades internacionales de humanidades digitales no captan bien la realidad internacional (ver la muy comentada infografía de Terras basada en datos de centerNet)⁵⁷. Dacos sugiere que es un campo gobernado por Europa y Norteamérica, sobre todo Reino Unido y sus colonias, y propone una serie de medidas para equilibrar la representación para los que no 'juegan en casa', entre ellos la discriminación positiva, el uso del 'Globish', una adaptación del inglés para no-nativos y el uso de 'tarjetas blancas' para los nativos que hablan rápido o usan lenguaje inaccesible durante sus presentaciones.

Un área donde se perciben claramente tensiones en el actual modelo internacional es en el tema de las propuestas para congresos, ya analizado por Clavert y Dacos, y que combina aspectos lingüísticos (la dominación actual de un idioma, el inglés, y las desventajas que esto puede suponer para los no nativos) y culturales (cómo escribir una propuesta o un artículo; cómo evaluarla; los códigos socio-culturales inherentes en cualquier proceso de arbitraje; los distintos significados culturales de la participación en congresos; los distintos mecanismos de acreditación a los que están ligados, etcétera).

Un escollo central para unas humanidades digitales verdaderamente internacionales es la falta de visibilidad de su presente y de su pasado. Esto tiene varias facetas: una falta de reconocimiento de publicaciones ya existentes (Fiormonte, 2012: 60); una falta de estudios sobre su historia en varios países, tan importantes como referencia; (en muchas lenguas) una falta de publicaciones sobre el campo; y una falta de representación en las publicaciones internacionales en inglés. Es de esperar que, en el futuro, los volúmenes y compendios que pretenden ser 'representativos' del campo a nivel internacional presten atención a la representatividad geográfica, por ejemplo: hasta ahora una proporción muy alta de estas publicaciones ha estado dirigida por y para investigadores ingleses nativos; sobre todo, en Norteamérica.

⁵⁵ <<http://www.clavert.net/the-digital-humanities-multicultural-revolution-did-not-happen-yet/>> [22/11/2013].

⁵⁶ Tampoco hay que menospreciar el efecto del apoyo financiero institucional. De las 589 propuestas para el congreso de DH2014, 333 (el 56,5%) provienen de cuatro países históricos en términos de financiación: Estados Unidos, Alemania, Reino Unido y Canadá (Vanhouette, 2013).

⁵⁷ <<http://www.ucl.ac.uk/infostudies/melissa-terras/DigitalHumanitiesInfographic.pdf>> [22/11/2013].

Cabe resaltar el contraste que se da aquí entre la visibilidad en las publicaciones formales y la que existe en medios sociales, blogs u otros modos informales de comunicación o difusión. Clavert identifica la falta de visibilidad internacional del campo en francés, no obstante su historia y el gran interés en el campo en países francófonos⁵⁸, un fenómeno familiar en muchos otros países. Esto puede tener muchas explicaciones –algunas socio-económicas generales, otras vinculadas a la cultura académica de humanidades digitales– pero un motivo importante puede ser lo que Dacos llama las diferencias culturales en ‘egoreferencement’ (en francés) o ‘autoreferencia’ (Dacos, 2012: 12), muy notables en los medios sociales, aunque es probable que esos mismos medios promuevan transformaciones sociales en ese mismo sentido que tenderán a equilibrar las diferencias a largo plazo.

A lo largo de los últimos años, varios investigadores habíamos notado mucha actividad asociada a las humanidades digitales que o bien se producía en Hispanoamérica/el estado español o bien operaba dentro de la cultura, historia e idiomas de esos territorios (en parte cubierta por lo que en Europa se llama la filología hispánica, aunque su alcance era mucho más general). Pero esta actividad rara vez era conocida en el mundo de las humanidades digitales (dominado hasta ahora por la comunidad angloparlante), y difícilmente encontraba cauce oficial en otros campos disciplinares. Hasta hace poco era común oír que muchos investigadores en humanidades digitales que operaban en castellano no conocían la investigación de otros colegas que estaban trabajando en la misma área, y era difícil conocer la historia de humanidades digitales en castellano (en España e Hispanoamérica) por falta de foros comunes establecidos.

Fue por ese motivo que varios investigadores que trabajábamos en humanidades digitales en castellano o en portugués decidimos, en junio de 2012, organizar un Día de humanidades digitales/Día de humanidades digitais, organizado por varias entidades internacionales⁵⁹ y hospedado por UNAM⁶⁰.

Basado en el *Day of Digital Humanities*⁶¹, un evento internacional que pretende documentar el campo de las humanidades digitales durante 24 horas y responder a la pregunta ‘¿Qué es lo que hacen los humanistas digitales?’ (Rockwell *et al.*, 2012), esta propuesta pretendía ofrecer a los investigadores de humanidades digitales (en sentido amplio) un foro para hacer conocer su investigación, y para conocer el trabajo de otros. El resultado fue una demostración clara del grado de interés en humanidades digitales en el mundo hispanohablante y lusófono, y resultó en varias otras iniciativas como el *MapaHD*, que pretende responder a la pregunta ¿Dónde están, quiénes son y qué hacen los humanistas digitales⁶²? Esta iniciativa no reemplaza la necesidad de estructuras más formales, pero supone un buen complemento, y una manera de fomentar conversaciones, relaciones y acciones a niveles más informales de la actividad académica.

⁵⁸ Nos informa, por ejemplo, de la suscripción de más de 500 personas a la lista de humanidades digitales en francés en las primeras semanas de su funcionamiento (2012).

⁵⁹ centerNet, Humanidades Digitales Hispánicas, Sociedad Internacional (HDH), Facultad de Filosofía y Letras, Universidad Nacional Autónoma de México, Red de Humanidades Digitales (RedHD), Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa, Humanidades Digitais, Universidade de São Paulo. <<http://dhd2013.filos.unam.mx/>> [22/11/2013].

⁶⁰ Un artículo a fondo sobre esta experiencia está en vías de preparación.

⁶¹ <<http://digitalhumanities.org/centernet/initiatives/>> [22/11/2013].

⁶² <<http://mapahd.org/>> [22/11/2013].

Fue en parte como reacción a críticas vertidas por comentaristas como Fioromonte y Dacos que en enero de 2013 se lanzó la iniciativa *Perspectivas Globales: Humanística Digital (GO::DH)*, un Grupo de Interés Especial de ADHO que pretende “ayudar a romper las barreras que dificultan la comunicación y la colaboración entre los investigadores y estudiantes de las artes digitales, humanística y sectores de patrimonio cultural en las diferentes economías del mundo”⁶³. En sus pocos meses de existencia, GO::DH ha promovido varias iniciativas, para fomentar la traducción colaborativa de textos digital-humanísticos⁶⁴, para servir de presentación a neófitos⁶⁵, y para unir a investigadores interesados en la problemática de la representación global como tema científico⁶⁶. Esto puede ayudar a corregir tendencias estructurales en el campo a nivel internacional, pero no pueden funcionar sin expresiones claras y visibles del campo a nivel local, que vayan creando instrumentos de coherencia, tanto a nivel académico como a nivel práctico. No debe ser desconocida la impresión de Baraibar, quien afirma que “el problema, al menos en España, está más en la falta de una oficialidad de las Humanidades Digitales que en la ausencia de iniciativas”⁶⁷.

La aparición en el mundo castellanoparlante de dos asociaciones (Red de Humanidades Digitales/Red HD y Humanidades Digitales Hispánicas/HDH) y varias iniciativas (THATCamp Caribe⁶⁸, THATCamp Buenos Aires⁶⁹) en los últimos años es prueba de un vuelco importante en la representación, que promete crear los instrumentos necesarios para el desarrollo de unas humanidades digitales, empezando por el debate que desvelará las particularidades locales y regionales. Y es importante observar que de la misma manera que las humanidades digitales deben evitar e incluso criticar argumentos homogeneizantes en los materiales estudiados, también deben aplicar este paradigma a sus propios métodos y discursos. Este análisis debe integrar estudio y crítica de los orígenes y tendencias culturales de las herramientas que creamos y utilizamos: hasta ahora, buena parte de la historia y los conocimientos que tenemos de las TIC tienen un sesgo anglófono (lingüístico y cultural), como afirma Folaron (2012).

CONCLUSIONES

Como hemos visto, las humanidades digitales no representan una categoría fija de cuestiones humanísticas ni un juego de herramientas o metodologías estables. El mismo dinamismo, proporcionado por las características del medio digital, provoca inestabilidad e incertidumbre sobre su propio futuro. Pero es dudoso que estos debates sobre si es una disciplina propia o una actividad transdisciplinar, interdisciplinar, etcétera se resuelvan pronto, lo cual es una muestra de que se trata de un debate más profundo sobre la naturaleza actual de la ciencia en general. Este debate tiene un carácter especialmente intenso en

⁶³ <<http://www.globaloutlookdh.org/nosotros-2/?lang=es>> [22/11/2013].

⁶⁴ <<http://www.globaloutlookdh.org/translation-commons/>> [22/11/2013].

⁶⁵ <<http://www.globaloutlookdh.org/working-groups/491-2/>> [22/11/2013].

⁶⁶ <<http://www.humanidadesdigitales.net/index.php/encuentro2014/encuentro2014-en>> [22/11/2013].

⁶⁷ <<http://dhd2013.filos.unam.mx/porvistadeojos/2013/05/20/buenos-tiempos-para-las-humanidades-digitales-en-espanol/>> [22/11/2013].

⁶⁸ <<http://caribbean2013.thatcamp.org/>> [22/11/2013].

⁶⁹ <<http://buenosaires2013.thatcamp.org/>> [22/11/2013].

las humanidades, donde el papel de las humanidades digitales frecuentemente asume un carácter político (la existencia de al menos tres ‘manifiestos’⁷⁰ bajo su nombre no es un accidente) y donde el campo puede suscitar valoraciones desde polos opuestos: para algunos sería ‘el salvador’ de las humanidades, mientras para otros, sirve como ‘caballo de Troya’ para actores hostiles a la perspectiva crítica imperante en las humanidades⁷¹.

Alan Liu responde en negativo a su propia pregunta sobre si las humanidades digitales están preparadas para asumir su responsabilidad para representar las humanidades y la educación superior en un momento en el que estas últimas negocian una nueva relación con la sociedad posindustrial (2012). Según Liu, la tradición del campo proveniente de la informática humanística carece de perspectiva crítica en el plano cultural, mientras que la rama vinculada a los estudios de la comunicación peca de ejercer una crítica indiscriminada hacia la sociedad y el imperio global de la información. Iniciativas como *New Humanities*⁷² y *4Humanities* intentan saldar este déficit político y crítico, ofreciendo defensas de las humanidades dentro de un debate activo sobre su futuro.

La cuestión sobre cómo reimaginar las humanidades nos hace volver a las cuestiones fundamentales de las humanidades –analizar, criticar y comunicar la relación entre el ser humano, su historia, sus lenguas y sus culturas– en un momento en el que chocan contra la hibridización de los medios y las formas de comunicación; modelos de proceso de investigación e interacción humana; conceptos de autor y lector/usuario; y nuevas maneras de crear conocimiento y nuevos agentes para captar, preservar y difundirlo.

Es una oportunidad, como reconoce Megías, de cruzar una brecha que se abre entra la sociedad y las humanidades (2012: 14), devolviendo a las humanidades cierta agencia y relevancia social. Muchos han observado su alejamiento de los procesos de la creación y difusión del conocimiento (Numerico, Fiorimonte y Tomasi, 2010: 8) y este momento histórico puede suponer para las humanidades digitales un papel de mediación, entre las nuevas fuentes y formas de visualizar el conocimiento, la necesidad cada vez más apremiante de una perspectiva crítica sobre las herramientas que usamos (digitales o no), “la ‘humanización’ de la tecnología” (Rodríguez-Yunta, 2012) y la transferencia –incompleta, descontextualizada y que necesita reflexión humana– de parte de nuestra interacción humana (incluyendo los frutos de ella) al medio digital. Las humanidades digitales, ya no solamente en la periferia, ofrecen un canal importante para debatir e innovar en el espacio entra tecnología, sociedad y perspectiva crítica.

BIBLIOGRAFÍA

Anderson, Benedict, *Imagined Communities*, London, Verso, 1991.

Bertrand, Paul, “Les *digital humanities* sont-elles solubles dans le Steampunk?: Conférence inaugurale”, en *THATCamp Paris 2012: Non-actes de la non-conférence des humanités*

⁷⁰ *Digital Humanities Manifesto 2.0*, junio de 2009 <http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf>, *Manifesto for the Digital Humanities*, marzo de 2011 <<http://tcp.hypotheses.org/411>> y *Young Researchers in Digital Humanities: A Manifesto* <<http://dhdhi.hypotheses.org/1855>> [22/11/2013].

⁷¹ Ver, por ejemplo, los comentarios de Michael J. Kramer en <<http://www.michaeljkramer.net/issuesindigitalhistory/blog/?p=1221>> [22/11/2013].

⁷² <<http://www.newhumanities.org/>> [22/11/2013].

- numérique*, Paris, Éditions de la Maison des sciences de l'homme, 2012. Web <<http://books.openedition.org/editionsmsmh/327>> [22/11/2013].
- Borgman, Christine, "The Digital Future is Now: A Call to Action for the Humanities", en *Digital Humanities Quarterly*, vol. 3, nº 4 (2009), <<http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html>> [22/11/2013].
- Borràs, Laura, "The Internet as Interzone", ponencia en congreso *Charting the Interzone: a conference*, 2011, <<http://www.unibg.it/dati/bacheca/676/54571.pdf>> [22/11/2013].
- Bradley, John, "What the Developer Saw: an Outsider's View of Annotation, Interpretation and Scholarship", en *Digital Studies/Le champ numérique*, vol. 1, nº 1 (2009), <http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/143/202> [22/11/2013].
- Clavert, Frédéric, "The Digital Humanities multicultural revolution did not happen yet", entrada de blog, 26 de abril de 2013, <<http://www.clavert.net/the-digital-humanities-multicultural-revolution-did-not-happen-yet/>> [22/11/2013].
- Crymble, Adam y Flanders, Julia, "FairCite", en *Digital Humanities Quarterly*, vol. 7, nº 2 (2013), <<http://digitalhumanities.org/dhq/vol/7/2/000164/000164.html>> [22/11/2013].
- Dacos, Marin, "La stratégie du Sauna finlandais", en *Blogo Numericus*, mayo de 2013, <<http://blog.homo-numericus.net/article11138.html>> [22/11/2013].
- Deegan, Marilyn y McCarty, Willard (eds.), *Collaborative Research in the Digital Humanities*, Farnham/Burlington, Ashgate Press, 2012.
- Fiormonte, Domenico, "Towards a Cultural Critique of the Digital Humanities", en 'Controversies around the Digital Humanities', 2012, A special issue from *Historical Social Research/Historische Sozialforschung*, vol. 37, nº 3, <http://www.cceh.uni-koeln.de/files/Fiormonte_final.pdf> [22/11/2013].
- Flanders, Julia, "The Productive Unease of 21st-century Digital Scholarship", en *Digital Humanities Quarterly*, vol. 3 nº 3 (2009), <<http://digitalhumanities.org/dhq/vol/3/3/000055/000055.html>> [22/11/2013].
- Folaron, Deborah, "Translation, Globalization and Communication Technology", en *Translation Spaces*, 2012, pp. 5-31, <<http://benjamins.com/#catalog/journals/ts.1.02fol/sample>> [22/11/2013].
- Galina, Isabel, "¿Qué son las Humanidades Digitales?", en *Revista Digital Universitaria*, 1 de julio de 2011, <<http://www.revista.unam.mx/vol.12/num7/art68/>> [22/11/2013].
- Galina, Isabel, "Is There Anybody Out There? Building a global Digital Humanities Community", entrada en *Blog de RedHD*, 19 de julio de 2013, <<http://humanidadesdigitales.net/blog/2013/07/19/is-there-anybody-out-there-building-a-global-digital-humanities-community/>> [22/11/2013].
- Galina, Isabel, "Las Humanidades Digitales globales", entrada en *Blog de RedHD*, 8 de noviembre de 2013, <<http://humanidadesdigitales.net/blog/2013/11/08/las-humanidades-digitales-globales/>> [22/11/2013].
- Galison, Peter, *Image and Logic: A Material Culture of Microphysics*, Chicago, University of Chicago Press, 1997.
- Gibbs, Fred, "Critical Discourse in Digital Humanities", en *Journal of Digital Humanities*, vol. 1, nº 1 (2011), <<http://journalofdigitalhumanities.org/1-1/critical-discourse-in-digital-humanities-by-fred-gibbs/>> [22/11/2013].
- Kirschenbaum, Matthew, "What Is Digital Humanities and What's It Doing in English Departments?", en *ADE Bulletin*, 2010, <http://mkirschenbaum.files.wordpress.com/2011/01/kirschenbaum_ade150.pdf> [22/11/2013].
- Isasi, Carmen, "Edición digital. Retos nuevos en los nuevos recursos", en Mariña Arbor y Antonio F. Guadianes (eds.), *Estudos de edición crítica e lírica galego portuguesa, Verba*. Anexo 67, Universidad de Santiago de Compostela, 2010, pp. 353-368.

- Le Deuff, Olivier, "Humanisme numériques et littératies", en *Semen*, nº 34 (2012), pp. 117-134, <http://hal.archivesouvertes.fr/docs/00/77/68/05/PDF/LE_DEUFF_BON_SEH.pdf> [22/11/2013].
- Liu, Alan, "Where is Cultural Criticism in the Digital Humanities?", ponencia de 2011, <<http://liu.english.ucsb.edu/where-is-cultural-criticism-in-the-digital-humanities/>>, luego ampliado y publicado en *Debates in the Digital Humanities*, ed. Matthew K. Gold, Minneapolis, University of Minnesota Press, 2012 <<http://dhdebates.gc.cuny.edu/debates/text/20>> [22/11/2013].
- Liu, Alan, "The state of the digital humanities: A report and a critique", en *Arts and Humanities in Higher Education*, vol. 11 (2012), pp. 8-41.
- López Poza, Sagrario, "Humanidades digitales hispánicas", *Cincuentenario de la AIH*, A Coruña, 2014 (en prensa).
- Lucía, José Manuel, *El elogio del texto*, Madrid, Fórcola, 2012.
- McCarty, Willard, "Humanities computing as interdiscipline", 1999, <<http://www.iath.virginia.edu/hcs/mccarty.html>> [22/11/2013].
- McCarty, Willard, "Getting there from here: Remembering the future of digital humanities", en *LLC: the Journal of Digital Scholarship in the Humanities* (en prensa). Basado en Busa Award Lecture 2013, <<http://www.youtube.com/watch?v=nTHa1rDR680>> [22/11/2013].
- Moulin, Claudine et al., "Research Infrastructures in the Digital Humanities", *Science Policy Briefing 42*, Strasbourg, European Science Foundation, 2011, <http://www.esf.org/file-admin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf> [22/11/2013].
- Nagasaki, Kiyonri y Muller, Charles, "Trends of Digital Scholarship in the Humanities in Japan", *Presentación a congreso aaDH*, 2012, <<http://aa-dh.org/images/2011/12/DHA2012-abstracts1.pdf>> [22/11/2013].
- Numerico, Teresa; Fiormonte, Domenico y Tomasi, Francesca, *L'umanista digitale*, Bologna, Il Mulino, 2010.
- Nyhan, Julianne; Flinn, Andrew y Welsh, Anne, "A short Introduction to the Hidden Histories project and interviews", en *Digital Humanities Quarterly*, vol. 6, nº 3 (2012), <<http://digitalhumanities.org/dhq/vol/6/3/000130/000130.html>> [22/11/2013].
- Prescott, Andrew, "Small worlds and big tents", entrada de blog de 5 de mayo de 2012, <<http://digitalriffs.blogspot.co.uk/2013/05/small-worlds-and-big-tents.html>> [22/11/2013].
- Ramsay, Stephen, "DH Types One and Two", entrada de blog de 3 de mayo de 2012, <<http://stephenramsay.us/2013/05/03/dh-one-and-two/>> [22/11/2013].
- Rockwell, Geoffrey, "Humanities Computing Challenges", entrada de blog en *Theoreti*, August 31, 2004, <<http://theoreti.ca/?p=544>> [22/11/2013].
- Rockwell, Geoffrey et al., "The Design of an International Social Media Event: A Day in the Life of the Digital Humanities", en *Digital Humanities Quarterly*, vol. 6, nº 2 (2012), <<http://www.digitalhumanities.org/dhq/vol/6/2/000123/000123.html>> [22/11/2013].
- Rodríguez-Yunta, Luis, "Las humanidades digitales, ¿una mera etiqueta o un campo por el que deben apostar las ciencias de la documentación?", en *Notas ThinkEPI 2013*, 12 de septiembre de 2012, <<http://www.thinkepi.net/humanidades-digitales-etiqueta-campo-apostar-ciencias-documentacion>> [22/11/2013].
- Schreibman, Susan; Siemens, Ray y Unsworth, John (eds.), *A Companion to Digital Humanities*, Oxford, Blackwell, 2004, <<http://www.digitalhumanities.org/companion/>> [22/11/2013].
- Suárez, Juan Luis, "El Humanista Digital", en *Revista de Occidente*, nº 380 (2013), pp. 5-22, <http://www.cultureplex.ca/media/publications/JLS_El_Humanista_Digital.pdf> [22/11/2013].
- Svensson, Patrik, "Humanities computing as digital humanities", en *Digital Humanities Quarterly*, vol. 3, nº 3 (2009), <<http://digitalhumanities.org/dhq/vol/3/3/000065/000065.html>> [22/11/2013].

- Svensson, Patrik, "The Landscape of Digital Humanities", en *Digital Humanities Quarterly*, vol. 4, nº 1 (2010), <<http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>> [22/11/2013].
- Svensson, Patrik, "Envisioning the Digital Humanities", en *Digital Humanities Quarterly*, vol. 6, nº 1 (2012), <<http://digitalhumanities.org:8080/dhq/vol/6/1/000112/000112.html>> [22/11/2013].
- Thaller, Manfred (ed.), "Controversies around the Digital Humanities", A special issue from *Historical Social Research/Historische Sozialforschung*, vol. 37, nº 3 (2012).
- Vanhoutte, Edward, "The World of Digital Humanities: Digital Humanities in the World", ponencia en *Cross Country/Faculty Workshop on Digital Humanities: Prospects and Proposals*, North-West University Potchefstroomkampus, South-Africa, 13 de noviembre de 2013, <<http://www.slideshare.net/edwardvanhoutte1/the-world-of-digital-humanities-digital-humanities-in-the-world>> [22/11/2013].



RESUMEN

El aumento de instituciones e iniciativas en humanidades digitales en los últimos años no ha ayudado a definir el campo hasta el momento, y la diversidad en las etiquetas empleadas para describir la actividad ('humanidades digitales', 'informática umanística e la cultura digitale', 'humanités numériques') no solamente refleja circunstancias locales distintas, sino divergencias importantes en la manera de concebirlo. ¿Se puede hablar de las humanidades digitales como un fenómeno único a nivel internacional, o más bien representa una acumulación de prácticas y perspectivas interrelacionadas? Este capítulo estudiará los centros y límites de un campo que se transforma con relativa rapidez, y señala algunos de los retos que le esperan todavía en medio de transformaciones académicas y sociales significantes causadas por la era digital.

El brote de nuevas asociaciones internacionales en el campo ha puesto de relieve, además, la gran cantidad de investigación digital que toma lugar fuera de los foros internacionales, donde los protagonistas anglófonos juegan con cierta ventaja. Con iniciativas como Humanidades Digitales Hispánicas (HDH), Red de Humanidades Digitales (RedHD) y el Día de Humanidades Digitales 2013 (DíaHD2013), se ha pretendido visualizar y fomentar una tradición ya con raíces profundas, pero no siempre muy bien conocida, en países castellanoparlantes, y esto entabla retos para la visibilidad que forman parte de un debate importante sobre cómo conseguir unas humanidades digitales con un carácter auténticamente global.

Palabras clave: Humanidades digitales, informática humanística, erudición digital, perspectiva global.

ABSTRACT

The growth in institutions and initiatives in digital humanities in the last few years has yet not led to greater definition in the field, and the different labels used for the activity ('humanidades digitales', 'informática umanística e la cultura digitale', 'humanités numériques') not only reflect differing local circumstances, but also important divergence in the manner of conception. Can we talk of the digital humanities as a unified phenomenon on an international level, or do they rather represent an accumulation of interrelated practices and perspectives? This chapter will study the centres and limits of a field which experiences continuous and rapid change, and aims to highlight some of the key challenges which await the digital humanities in a period of major academic and social transformation caused by the digital age.

Meanwhile, the rapid evolution of new international associations in the field has highlighted the large amount of digital research carried out away from the glare of interna-

tional fora for digital humanities, where Anglophone academics have a certain advantage. Spanish-language initiatives like Humanidades Digitales Hispánicas (HDH), Red de Humanidades Digitales (RedHD) and el Día de Humanidades Digitales 2013 (DíaHD2013, which also has a Portuguese language component) have attempted to improve the visibility of a tradition which already has strong roots, but which does not always figure prominently on the international stage, and these form part of a broader ongoing debate about how to ensure that digital humanities become a truly international phenomenon.

Keywords: Digital humanities, humanities computing, digital scholarship, global outlook.

Proyectos sociales y de aula a partir del aprendizaje colaborativo virtual. Experiencias desde la tutoría en Diplomado de AUSJAL¹

DIEGO AGUDELO GRAJALES²
Pontificia Universidad Javeriana, Cali
dagudelo@javerianacali.edu.co

INTRODUCCIÓN

En contextos de globalización y de tecnologías de información y comunicación (TIC) que impactan desde sus desarrollos y aplicaciones el mundo contemporáneo son muchos los lenguajes comunes que se van tejiendo y que dan por hechos las mismas concepciones y posibilidades, dando lugar por ejemplo a identificaciones que son en sí mismas diversas como es el aprender *de* la tecnología y el aprender *con* la tecnología (Montes-Ochoa, 2006).

La diferencia en estas dos expresiones está en las dos preposiciones que las distinguen y que dan lugar a campos de significación distintos, el primero enfatiza en las posibilidades de las TIC como herramientas para transmitir información al estudiante a bajo costo y en tiempos mínimos con la posibilidad de acceder a información diversa y con la posibilidad de comparar sus posibilidades en relación al objetivo que se propone (Marti, 2003). Este énfasis, sin embargo, no tiene en cuenta el papel del estudiante en la apropiación de la tecnología, ni su conocimiento previo y menos sus intereses. Por su parte, el segundo, es decir el aprender *con* la tecnología considera a las TIC como herramientas para la construcción de conocimiento, por lo que supone que las TIC tienen sentido en la medida en que los actores del hecho educativo las usen para este fin (Liaw, 2004) e incluso para negociar el sentido de sus observaciones en contextos colaborativos (Karagiorgi y Symeou, 2005).

Cada uno de estos modelos han generado estudios privilegiando algunos factores específicos, por ejemplo es común en nuestros países de América Latina,

¹ Ponencia en el evento: *Humanidades Digitales: desafíos, logros y perspectivas de futuro. I Congreso Internacional de la HDH: Humanidades Digitales Hispánicas. Sociedad Internacional, A Coruña (España), 9-12 de julio de 2013.*

² Profesor a tiempo completo en Pontificia Universidad Javeriana Cali, Tutor del Diplomado Ausjal, Director del grupo de investigación De Humanitate, PHD en Teología.

que los gobiernos tomen como una bandera política hacer un gran esfuerzo por mostrar la importancia del modelo de *aprender de la tecnología*, pues es un modelo que puede mostrar indicadores altos de cumplimiento y por tanto, cierta facturación de orden político, porque cada entrega de equipos de cómputo es un indicador más que se está cumpliendo. Por ejemplo en Colombia, el Estado hace una gran inversión en computadores para toda la red pública de educación, y efectivamente estos equipos se entregan a las comunidades escolares con la asistencia de la prensa nacional para difundir las buenas obras y el cumplimiento de una de las metas del gobierno, pero como no existen las condiciones de seguridad para conservarlos o las redes para habilitarlas, estos equipos en muchos lugares de la geografía colombiana no existen, aunque ya estén registrados por parte del Estado como entregados y por tanto las comunidades sólo quedaron con un espejismo. Es frecuente escuchar en las noticias que en la escuela x de una región apartada del país fueron hurtados los computadores que el Estado previamente había entregado.

En este mismo modelo, otro factor que se suele destacar es en relación a los beneficios del uso, pues los resultados muestran que los estudiantes disfrutaron el uso de las tecnologías en el salón de clase y tienen mayores posibilidades de acceso al conocimiento, con lo que se pretende cerrar la brecha entre ricos y población desfavorecida que en nuestros contextos es bien acentuada esta diferencia. No obstante, una limitación de estos estudios es que la satisfacción por sí sola no está ligada con la construcción de conocimiento, y que es necesario explorar si los estudiantes que disfrutaron más de la vinculación de las TIC en su educación, dedican más tiempo a aprender y usan recursos novedosos propios de las TIC para mejorar o para resolver problemas de comprensión.

Desde el modelo de *aprender con la tecnología* se encuentran algunos estudios que indagan acerca del proceso mismo de aprendizaje en contextos educativos enriquecidos con TIC. Estos estudios analizan en qué medida la integración de las TIC en la educación favorece el uso y promoción de habilidades cognitivas de alto orden tales como el pensamiento crítico, la resolución de problemas y la transferencia del conocimiento (Hou y Wu, 2011), así como el aprendizaje basado en proyectos (De Wever, Van Keer, Schellens y Valcke, 2010; Du y Xu, 2007), y la apropiación de los docentes de las TIC como herramientas para enriquecer los escenarios de aprendizaje-enseñanza (Ochoa y Montes, 2006), o el diseño de nuevos modelos formativos con soporte de tecnología para generar una oferta educativa de calidad en la escuela primaria y secundaria (Sánz, 2012).

Una característica importante de estos estudios es el seguimiento en el tiempo de los procesos de aprendizaje, bien sea a lo largo de una unidad académica dentro de una asignatura, en un semestre académico o a lo largo de una carrera profesional. Los resultados de estos estudios muestran que el uso de las TIC en ambientes colaborativos favorece la construcción de conocimiento y la resolución de problemas cognitivamente complejos. Las limitaciones que presentan este tipo de investigaciones es su larga duración, los costos y el tipo de análisis exhaustivo de un contexto educativo particular desde el cual se pretende generalizar, lo cual obviamente sirve como punto de referencia pero no de aplicación en las prácticas educativas y corresponde a cada comunidad académica la respectiva validación desde sus características específicas.

Estudiadas estas dos concepciones por parte de AUSJAL se decide por el aprendizaje con la tecnología y diseña desde la red de homólogos el Diplomado de AUSJAL. Este Diplomado: "Formación en Red de Docentes Universitarios en TIC"

(Primer Nombre), y luego, "Diseño de Proyectos Virtuales con Énfasis en Proyectos Sociales o Proyectos Académicos", son realizados desde una coordinación académica, técnica y administrativa que tienen la sede en la Pontificia Universidad Javeriana de Bogotá y se utiliza hasta la fecha (2013-2) la plataforma Black Board para este fin.

El propósito de este artículo es mostrar cuál ha sido la experiencia alrededor del desarrollo de estos Diplomados que han permitido diseñar, ejecutar y hacer seguimiento a una serie de proyectos que intentan contribuir en el mejoramiento de la calidad educativa y en la Responsabilidad Social Universitaria estudiando problemas sociales y buscando alternativas de solución que sean creativas y fruto del trabajo colaborativo. El desarrollo de este artículo se plantea en tres partes, una primera que corresponde a saber qué es la red AUSJAL, la segunda es el cómo se realiza o se ha realizado el Diplomado en las nueve cohortes que lleva hasta el momento; y la tercera, que corresponde a la selección de algunos proyectos que se han diseñado, los que han avanzado a una etapa de implementación.

Estos tres puntos corresponden a tres planos distintos que están interconectados en función de una pregunta orientadora ¿Es posible contribuir en forma significativa en la transformación social desde la formación virtual? El primero se refiere a las exigencias de trabajar en Red por parte de las instituciones universitarias de América Latina que comparten una identidad, una pedagogía, un contexto y unas fuentes. La segunda parte se referirá a la relación que la universidad establece con el conocimiento y con la sociedad, a través de las tecnologías de la información y la comunicación. Por último, se responde a la pregunta fundamental de esta presentación en relación a la contribución desde la virtualidad a la transformación de la sociedad. El lugar de análisis es desde la experiencia como tutor virtual de varias cohortes en las que se adquieren nuevas habilidades pedagógicas pero ante todo, ciertas actitudes que permitan tener la mente abierta a lo imprevisible, nuevas formas de interacción o la comprensión de nuevos significados aportados por los estudiantes, entre otros.

1. RED AUSJAL

AUSJAL es una red universitaria que articula a 30 universidades e instituciones de educación superior confiadas a la Compañía de Jesús en 14 países de América Latina. Ella como red tiene a su vez *Redes de homólogos* entendidas como equipos de profesores, profesionales o administradores que trabajan en forma conjunta en relación a proyectos específicos o estratégicos. Una de esas grandes estrategias es el fortalecimiento de las tecnologías de la comunicación y en especial de la virtualidad para favorecer en primer lugar la consolidación de esas redes académicas mediante cursos, diplomados virtuales. Para que esta estrategia sea sostenible se ofreció un Diplomado en Formación de tutores en red donde cada institución de la red facilitó la formación de un profesor, que luego tenía la responsabilidad de ser tutor del Diplomado orientado a la formación en virtualidad para diseñar proyectos sociales o de aula a través de la virtualidad. De esta manera la red de apoyo en la formación virtual va aumentando en número de participantes así como de iniciativas y proyectos sociales que pretenden contribuir desde el lugar específico de la educación para la Responsabilidad Social, el compromiso con los más desfavorecidos, la inclusión social, entre otros.

Desde un comienzo la Asociación en 1985 se buscó trabajar en red para trabajar en proyectos comunes y que desde allí se garantice su vigencia evitando caer en un escenario burocrático, tentación fácil y posible en este tipo de organizaciones, por eso la clave no estaría sólo en la voluntad de los rectores, sino en la creación y fortalecimiento de estas redes de homólogos con todo el respaldo institucional, por un lado, y al mismo tiempo, el incentivo de nuevas formas de trabajo colaborativo caracterizados por diseñar y ejecutar proyectos alineados con las estrategias megas de cada una de las instituciones participantes. Poco a poco, valorando positivamente, los avances de la Asociación, otras instituciones se han sumado a la red. Del mismo modo para este fortalecimiento como red ha sido el acuerdo la elaboración de un ideario común en materia de identidad y de los desafíos que la realidad social de los países latinoamericanos planteado a las universidades, en particular a las de AUSJAL (V Asamblea de Rectores en 1990, en Quito), que luego de algunos años de reflexión dio lugar el documento "Desafíos de América Latina y Propuesta Educativa de AUSJAL" (1995) avanzando en la definición de la Misión y Visión de la red en el contexto de las sociedades latinoamericanas.

Una vez fortalecida la Asociación desde una estructura de identidad se pasó a los diseños de Planes Estratégicos que es donde va a tomar forma los diferentes proyectos de las Redes de homólogos <http://www.ausjal.org/redes.html>, entre los cuales está el de las TIC: <http://www.ausjal.org/tics.html>.

Es en ese proyecto macro de formación en la TIC que surge el planteamiento de estos Diplomados, pero ¿a qué responde?

Algunos datos sobre América Latina pueden ilustrar desde las cifras el panorama general que se constituye en motivación para promover el encuentro en red de universidades para pensar sobre el quehacer como instituciones educativas en un contexto específico. De acuerdo a la estadística de la Comisión Económica para América Latina CEPAL el 12% de la población está en pobreza absoluta (70 millones de personas)³, es decir, 3 de cada 10 personas están en esa condición de extrema pobreza y esa situación irremediamente afecta todos los escenarios de socialización como educación, salud, seguridad que se expresa en fenómenos sociales muy complejos como violencia, drogadicción, incorporación de menores en grupos insurgentes como guerrillas y paramilitarismo.

Ahora bien, los rasgos característicos de esta situación de pobreza están asociados en América Latina con la desocupación, con el aumento de la población demográfica, con las bajas remuneraciones laborales o la inequidad en los salarios, disminución de los puestos de trabajo que sumados a otros hechos como el narcotráfico, la falta de escolaridad, la creciente inflación, la dependencia económica, entre otros, ponen en evidencia una situación generalizada de insatisfacción de las necesidades básicas y lo que resulta más preocupante, la falta de esperanza para tener perspectivas de mejoramiento, por más que los gobiernos se empeñen en mostrar cifras de crecimiento.

No es necesario en este punto sumar la inestabilidad política y la tendencia polarizada de la opinión pública del ejercicio de liderazgo de los políticos y de los discursos "mesiánicos" que suelen escucharse todos los días en nuestros países a través de los noticieros o de la forma como se percibe al Estado como débil y

³ http://estadisticas.cepal.org/cepalstat/WEB_CEPALSTAT/Library/CEPALSTAT/Infograficos_pdf/Inf_1_es.pdf [consultado 05/05/2013].

sin legitimidad pues no es garante de los derechos de sus ciudadanos. Pero lo que sí llama la atención es que este panorama contrasta con ese triunfalismo del consumo y los ritos de prosperidad y éxito que propone el mercado generando un cambio cultural en los hábitos sociales, en las representaciones que crea una sensación de libertad. El problema es que esta posibilidad es válida para unos pocos que tienen los medios para lograrla mientras las mayorías permanecen contemplativas de estas posibilidades (Lechner, 2000).

En esta caracterización de la sociedad las tecnologías de la información y comunicación han sido muy importantes en la configuración de nuevas formas de socialización. En efecto las TIC en la actual "sociedad del conocimiento", como la ha denominado la UNESCO (2005), son muy importantes para el desarrollo de la sociedad en cuanto son fuentes de bienestar y equidad en el acceso a los bienes y servicios necesarios que requieren las mayorías pobres para tener oportunidades. Esta es precisamente la motivación que tiene en las universidades latinoamericanas y especialmente, las universidades jesuitas, de presentar programas virtuales que contribuyan a responder con alternativas creativas de desarrollo a los problemas que tiene la sociedad.

La tendencia generalizada de las universidades es centrarse en programas de rápido retorno y de gran demanda social, por eso se concentra en áreas como administración, negocios, gestión que son muy importantes, pero que no deben ser los únicos, ni deben estar desarticulados de sus contextos que exigen alternativas. Paralelamente con ellos es importante incentivar otros programas que estén orientados hacia la gestión de lo social que atienda demandas del entorno en relación con la pobreza, la alfabetización, la mala nutrición, la violencia, entre otras, pero reconociendo que las universidades en forma separada difícilmente pueden responder a estas necesidades y por ello la exigencia de consolidar iniciativas como la Red Universitaria Interreligiosa de América Latina y el Caribe –REUNIR–, la RICESU, Red de Universidades Católicas de Educación Superior, o la que se tiene como referencia en este trabajo, la Red AUSJAL.

2. EL PROYECTO DE DIPLOMADOS VIRTUALES COMO UNA EXPRESIÓN DE LA RELACIÓN UNIVERSIDAD-CONOCIMIENTO, TIC-SOCIEDAD

Un poco tarde la universidad latinoamericana ha entendido que el desarrollo de las comunicaciones, la informática ha transformado las relaciones sociales, las formas de aprender y la realidad en su totalidad, por lo menos en la forma de comprenderla, y que esta lentitud en la respuesta hace que en cierto modo estén desfasadas las propuestas si se atiende a la actualización de los medios. Pero de todos modos, la universidad sigue manteniendo su lugar en la sociedad, así no sea en la vanguardia de los cambios tecnológicos, sí como institución crítica de la sociedad y como generadora de conocimiento que permita responder a problemas con sentido de responsabilidad y anticipación, considerando aspectos que tienen que ver con lo humano, con la dignidad, con la solidaridad.

Así, la universidad mantiene su lugar en la sociedad buscando ser más creativa en la mediación del conocimiento. Es lo que P. Levy llama cibercultura (1998), en la cual las tecnologías han modificado las capacidades cognitivas del ser humano y las nuevas formas de acceso a la información y al conocimiento y el desarrollo de habilidades y capacidades que permiten incrementar el potencial de los proyectos colectivos.

Además el modo de interacción en red crea vínculos, como lo describe P. Levy, como un componente del corpus intangible que forma la red informática mundial, *World Wide Web*, sin las consideraciones de tiempo, espacio, lugar, siempre en movimiento y transformación constante, exige de la universidad en primer lugar el diseño de currículos más flexibles e interdisciplinarios en función no de temas, sino de problemas que plantea el contexto; en segundo lugar, el medio a emplear para divulgar y producir conocimiento es a través del uso de tecnologías interactivas para la producción de conocimiento y con el propósito de responder a una exigencia de sostenibilidad y de alcance, éstos proyectos deben implementarse en Red. Ahora bien, todo esto es posible si además se mantiene activa la reflexión pedagógica como una especie de motor que dinamiza las nuevas formas de interacción y reconocimiento de los roles en proceso de aprendizaje por parte de los actores.

Esta es la opción tomada desde la Red de homólogos de AUSJAL propone, entre otros, el Diplomado titulado "Diseño de Proyectos Virtuales con énfasis en: proyectos de intervención social o proyectos de aula (DATIC)" que es el objeto de referencia en este estudio y que está pensado para brindar una formación virtual que responda a las necesidades del contexto de región y que su aproximación al conocimiento se hace a partir de una pedagogía específica como es la ignaciana. El Diseño del proyecto es el propósito principal del Diplomado y este debe tener las siguientes características: problemáticos (articulado con las necesidades del contexto), pertinentes (aporte alternativas de solución o tratamiento de las problemáticas), interdisciplinarios, internacionales (docentes de diferentes países y análisis de problemáticas semejantes entre los países), interuniversitarios (conformación de red en trabajo colaborativo en las etapas de formulación, diseño y ejecución), construcción en espiral (cada fase retoma la anterior y avanza hasta la definición del proyecto), pedagógicos, tecnológicos y competitivos.

La forma como está estructurado el Diplomado a partir de 4 módulos busca ir brindando los aportes pedagógicos y tecnológicos para el diseño del proyecto, que a su vez se constituyen en los tres ejes de cada Módulo. En el Módulo introductorio se hace un acercamiento general a la idea del proyecto y los objetivos globales. En el Módulo 1 se hace una contextualización de la educación virtual en el mundo actual y se estudian los principios de la pedagogía ignaciana para revisar la coherencia con la identidad de las instituciones. Igualmente se presenta una serie de problemáticas importantes de la realidad latinoamericana que requieren atención y se propone la metodología del marco lógico para definir los problemas y los objetivos.

En el Módulo 2 se estudia el aprendizaje colaborativo desde el enfoque histórico-cultural de Vygoyski y otros métodos como el aprendizaje basado en proyectos, problemas y casos. En referencia a lo tecnológico se presentan las diversas tecnologías de información y comunicación disponibles en el momento para la creación de ambientes virtuales de aprendizaje. Aquí el proyecto avanza hacia los requerimientos tecnológicos y su justificación pedagógica.

En el Módulo 3 el eje tecnológico se centra en lo que son los ambientes virtuales de aprendizaje y en el pedagógico en el debate sobre el aprendizaje como cambio conceptual y las características pedagógicas de los ambientes virtuales, lo que permite avanzar en el proyecto en el diseño de estos ambientes con sus énfasis pedagógicos. En el Módulo 4 se propone una validación y evaluación de los procesos de aprendizaje en ambientes virtuales, por eso en el eje pedagógico

se estudia la evaluación con sus procesos de validación e instrumentos para ser aplicados en el proyecto.

Cada uno de estos módulos está distribuido en unidades y éstas en temas, los cuales mediante "guías didácticas" se proponen cada uno de los temas en tres fases: una primera fase de "punto de partida" donde se busca hacer explícita o saber cuál es el aprendizaje o experiencia previa que se tiene del tema a tratar, luego la segunda fase es la "lectura virtual" que se propone como insumo o como contenido y una tercera fase titulada "Construyendo conocimientos" que consiste en articular lo aprendido en relación con la construcción del proyecto. La herramienta más usual para desarrollar estas propuestas son el foro sincrónico y especialmente los asincrónicos pues permite el análisis y la discusión colectiva, así como los trabajos colaborativos de análisis de casos por ejemplo. Esta metodología busca desarrollar entonces habilidades y competencias específicas que ayuden a la apropiación de estos nuevos saberes con herramientas concretas. Por otra parte, se pretende incidir en algunas de las áreas problemáticas para América Latina, a través de los proyectos de diseños virtuales que elaboren la red de homólogos.

El Diplomado tendrá en cuenta la propuesta pedagógica ignaciana, por ello considera el ciclo de aprendizaje que incluye aspectos como: contextualización, observación reflexiva, conceptualización, experimentación activa y evaluación; los cuáles complementarán la funcionalidad de la plataforma seleccionada.

Proceso de acompañamiento durante el desarrollo del Diplomado prevé un acompañamiento tecnológico y pedagógico. El Diplomado cuenta con un equipo de tutores especializados; quienes asesorarán los proyectos que surjan a partir de la participación de los docentes en las redes de homólogos. Brindarán retroalimentación y orientaciones específicas, para apoyar el aprendizaje en los procesos de acceso al conocimiento, por parte de los participantes en el Diplomado; principalmente, en lo referente a la apropiación tecnológica y al desarrollo de la propuesta pedagógica.

El Diplomado tiene una duración de cinco meses y tiene fechas de inicio y conclusión de cada módulo lo que quiere decir que su desarrollo va en forma sincrónica y con resultados en la finalización de cada Módulo en relación con los avances en el diseño de los proyectos. Se propone en el Módulo introductorio algunas áreas de trabajo dentro de las cuales se espera, se inscriban las problemáticas latinoamericanas que se constituyen en fuente de los proyectos. Estas áreas de trabajo son: Salud (nutrición, prevención, aguas, salud mental, salud pública...); Educación (políticas públicas, cubrimiento, mejoramiento de la calidad, multilingüismo...); Pobreza (pobreza extrema, distribución del ingreso); Derechos humanos (problemas de género, valores, justicia restaurativa, resiliencia); Empleo (Pymes, generadores de empresa, economía solidaria...); Vivienda social, servicios públicos; Incorporación de las tic como motor de desarrollo (sociedad de la información, transformaciones culturales, impactos en la sociedad, comunicación); Medio ambiente y desarrollo sostenible (sustentable); Democratización, gestión pública, ciudadanía, gestión social, transparencia, ética.

3. EXPERIENCIAS SIGNIFICATIVAS

En las primeras 8 cohortes hasta el años 2012, desde 2005-2006 cuando se desarrolló la primera cohorte se han desarrollado 81 proyectos de los cuales 45 de ellos están centrados en el área de educación en diferentes niveles y

el segundo es sobre Diseño y utilización de TIC con 9 proyectos, seguido del área de la salud (prevención y promoción) y la gestión ambiental cada una con 4 proyectos. Las demás áreas tienen entre uno o dos proyectos. Esta estadística se puede ver mejor en una ilustración de los anexos.

Destaco dos proyectos por su tiempo de desarrollo, ejecución y alcances. El primer proyecto es uno que está a punto de iniciar el 29 de julio de 2013 llamado "Diplomado Gerencia Social-Virtual" propuesto como Servicio de Extensión, propuesto en la última cohorte por dos profesores. Este diplomado está inscrito en la Facultad de Ciencias Económicas y Administrativas de la Pontificia Universidad Javeriana de Cali, particularmente del Departamento de Gestión de Organizaciones, en respuesta a la necesidad presente en las organizaciones de entender y reflexionar sobre la situación social en América Latina, a través del cual los estudiantes tendrán aproximaciones al concepto de la Gerencia Social y al quehacer de un gerente social, así mismo tendrán un acercamiento a las diferentes herramientas.

En respuesta a esta necesidad y las condiciones actuales de nuestro país y de nuestro continente, donde las sociedades se sacuden ante transformaciones socioeconómicas y políticas, la Gerencia Social es un campo de conocimiento y desempeño, que entendida como la Gerencia del Cambio, requiere una permanente construcción y una mirada flexible, holística y transformadora de la realidad social. La Gerencia Social surge como una disciplina donde convergen otras áreas de conocimiento, que la nutren y le facilitan precisamente su papel integrador, pues las habilidades gerenciales que desarrolla un Gerente Social, están fundamentadas en el reconocimiento de la dignidad humana y la comprensión del contexto mundial, regional y local.

El diplomado se ofrece en una modalidad 100% virtual, esto implica que las sesiones presenciales se transforman, dando paso a un mayor número de horas de trabajo independiente con un acompañamiento permanente de los tutores del diplomado. Además, tiene como propósito la comprensión, el significado y la práctica de la gerencia social en el mundo académico y profesional.

El diplomado propone una serie de actividades formativas que le permiten alcanzar los objetivos de aprendizaje. De manera general, estas actividades implican la lectura y revisión previa de recursos (tales como mapas conceptuales, videos y audios), que son fundamentales para su desarrollo, la explicitación del conocimiento que usted ha construido y la consolidación de sus aprendizajes.

Como objetivos se propone en primer lugar:

- Analizar y reflexionar sobre el surgimiento de la Gerencia Social como disciplina que genera una nueva forma de asumir la construcción del desarrollo social, desde las instancias: privada, pública o fundacional.
- Conocer y comprender el contexto de la pobreza. Conocer las diferencias y similitudes que existen en la dirección y conformación de organizaciones en los sectores privado, público y fundacional y las instancias de participación ciudadana.
- Conocer los programas sociales que adelantan fundaciones y empresas sociales y sus impactos en la comunidad
- Conocer las herramientas de la Gerencia Social, realizar un acercamiento al concepto de capital social y comprender la aplicación del mercadeo social en el campo social.

Lo interesante para destacar de esta propuesta no es el Diplomado en sí mismo, sino el contexto de desarrollo y los espacios de red y productos que está articulando. Es una propuesta innovadora que sirve como plataforma de encuentro entre gerentes sociales que suelen tener problemáticas comunes y que mediante este programa puede hacer análisis conjuntos y específicos al mismo tiempo sobre la pobreza y las formas o iniciativas cómo pueden enfrentarse para contribuir en la disminución de esa realidad. Así, una vez realizada la formación en el Diplomado por parte de los profesores Nazly Duque y Santiago Londoño, que tenían a cargo la asignatura Gerencia Social en el programa de Administración de Empresas, tomaron sus notas de clase e hicieron una indagación bibliográfica que les permitiera robustecer las líneas temáticas del programa de su asignatura o syllabus. Con estos dos insumos, además de los estudios de caso y sistematización de consultorías e este nivel en organizaciones sociales se planteó por parte de la profesora Nazly escribir un texto universitario titulado: "La Gerencia Social y sus nexos con el concepto de pobreza. Primer acercamiento" (PUJ, Cali, junio 2013), que será guía del curso presencial y del mismo Diplomado virtual.

Este proyecto ha sido presentado en México en la universidad Iberoamericana y se propone desarrollar en el 2014 un acercamiento desde la realidad de ese país a partir de un convenio interuniversitario. De igual forma este Diplomado se abre a toda la región latinoamericana y aunque está diseñado instruccionalmente en modalidad virtual está en su fase de implementación y poco a poco se va socializando en red para que vaya fortaleciéndose.

El otro proyecto que presento es "Profesionalización de docentes en ejercicio, a través de Comunidades de Aprendizaje B-Learning dirigida a los participantes del Centro de Profesionalización de Fe y Alegría" en Venezuela. La información, el contenido y las figuras es tomado de la presentación de Verónica Trenard, Directora del Centro de Profesionalización de Fe y Alegría y Coautora de este proyecto, en el III ENCUESTO RED EDUTIC-AUSJAL -Pontificia Universidad Javeriana- Bogotá el 16, 17 y 18 de mayo de 2012 en el cual se presenta la Retrospectiva y Avance en la implementación del modelo de profesionalización docente a partir de la virtualización (Fig. 1).

La intención de este proyecto es incrementar el índice de profesionalización de docentes en ejercicio para que puedan brindar un mejor servicio y adquieran o mejoren en el uso de didácticas que responda a la realidad socioeducativa específica y especialmente de las comunidades más deprimidas (zonas urbanas populares, zonas rurales, indígenas y afros), que es donde Fe y Alegría como organización ofrece sus servicios. Esto significa que gran parte de la población infantil de los sectores deprimidos tienen acceso a la educación pero ésta no cuentan con profesionales en la docencia y tampoco los docentes tienen posibilidades de acceder a la formación universitaria para profesionalizarse.

Antes de continuar en la presentación del proyecto es importante decir que "Fe y Alegría es un Movimiento de Educación Popular Integral y Promoción Social dirigido a los sectores empobrecidos y a los excluidos, potenciando su desarrollo personal y participación social" (<http://www.feyalegria.org/>). Como organización está en todos los países de América Latina (incluidos los países de Centroamérica), la República de Chad en el continente africano, así como en España e Italia en Europa. El desarrollo e influencia en todos los países es diferente, obviamente hay una semejanza por el contexto en América Latina y una diferencia grande respecto a África donde no hay otro contexto de comparación.

Retrospectiva y Avance de la Virtualidad en el CPFA

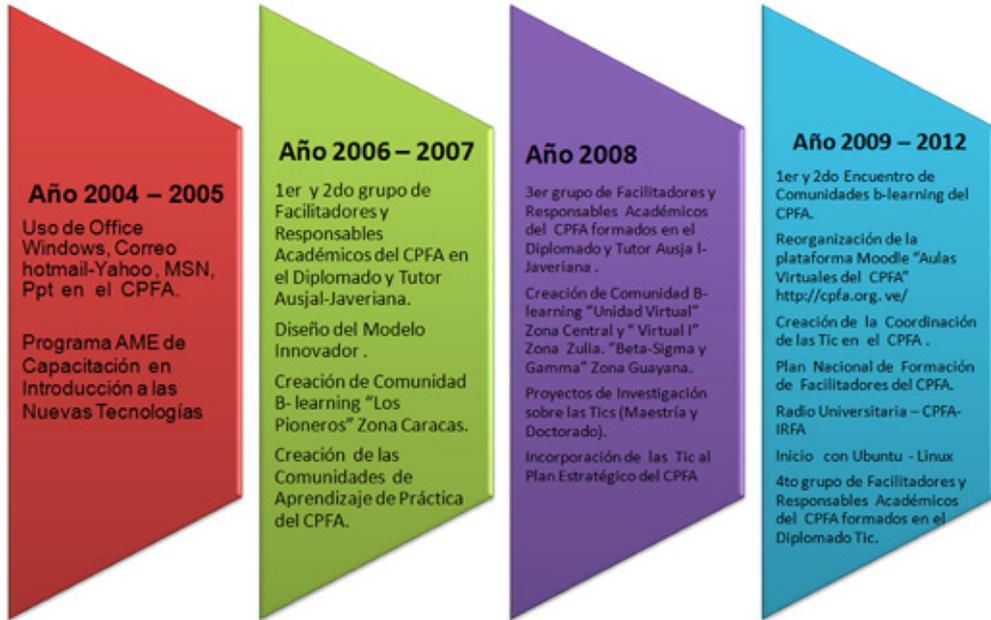


Fig. 1.

El modelo de profesionalización de docentes en ejercicio, que conduce a la obtención de licenciatura en educación, de la dirección nacional de educación universitaria de Fe y Alegría de Venezuela, implica el diseño e implementación de una propuesta académica, que permita la licenciatura de docentes en ejercicio no graduados y su formación permanente desde la práctica, a través de una estrategia de enseñanza-aprendizaje colaborativa: el B-learning, que combina las modalidades de educación presencial, educación virtual y el auto-aprendizaje, teniendo como componente transversal el uso de las TIC (<http://aulasvirtualescpfa.cpfa.org.ve/moodle/>).

La comunidad de aprendizaje B-learning tendrá la posibilidad de desarrollar una formación flexible, que le permita a los participantes y facilitadores, incorporar un nuevo paradigma educativo, en donde el trabajo colaborativo, la participación, el diseño, la innovación, la investigación aplicada y el sentido crítico, permitirá aplicar y validar los conocimientos alcanzados.

La organización de la población en este proyecto está organizada en siete comunidades de aprendizaje integradas por quince participantes en distintos espacios. Estas siete comunidades corresponden a las siete regiones en que se organiza el Centro de Profesionalización de Fe y Alegría (CPFA) en Venezuela.

En la primera etapa del proyecto se trabaja con una comunidad por región: Creación de ambiente de aprendizaje B-Learning. La Matriz Curricular está organizada en Fases, Ámbitos, Dimensiones, Áreas del Conocimiento y Ejes Transversales. Las Fases son los diferentes momentos que vive el participante en su

proceso de formación, las cuales son: Fase Introdutoria, Fase de Escolaridad y Fase Profesional o de especialidad (ver Fig. 2).

1. La fase Introdutoria diseñada para desarrollarse en un tiempo de 6 meses está conformada por cuatro bloques: I. Integración, II. Metodológico, III. Curricular, IV Proposición.
2. La fase de Escolaridad tiene dos ámbitos uno general: centrado en las dimensiones del "Aprender a aprender" y la "formación humana", y otro pedagógico centrado en la construcción de Proyectos de Aprendizaje a partir de los seminarios de: Comunicación y expresión, Pensamiento Lógico-Matemático, TIC, Metodología de Proyectos e Investigación Cualitativa (<http://aulasvirtualescpfa.cpfa.org.ve/moodle/>).
3. La fase de Especialidad está centrada en el ámbito de la Especialidad que implica la dimensión de formación en la construcción de la Especialidad, una propuesta de Innovación, Didácticas Especiales, una Matriz Personalizada, Rol Profesional (reflexionado a la luz del Perfil del Egresado). Reflexión de su proceso.

La matriz general de este modelo de licenciatura en educación puede verse en la figura 3 (Ejes transversales), en el cual se destacan las dimensiones de formación: humana, socio-política, proceso de enseñanza-aprendizaje, sistema educativo, "transformarnos para transformar espacios educativos" y el de la gestión educativa.

Fases del Diseño



Fig. 2.

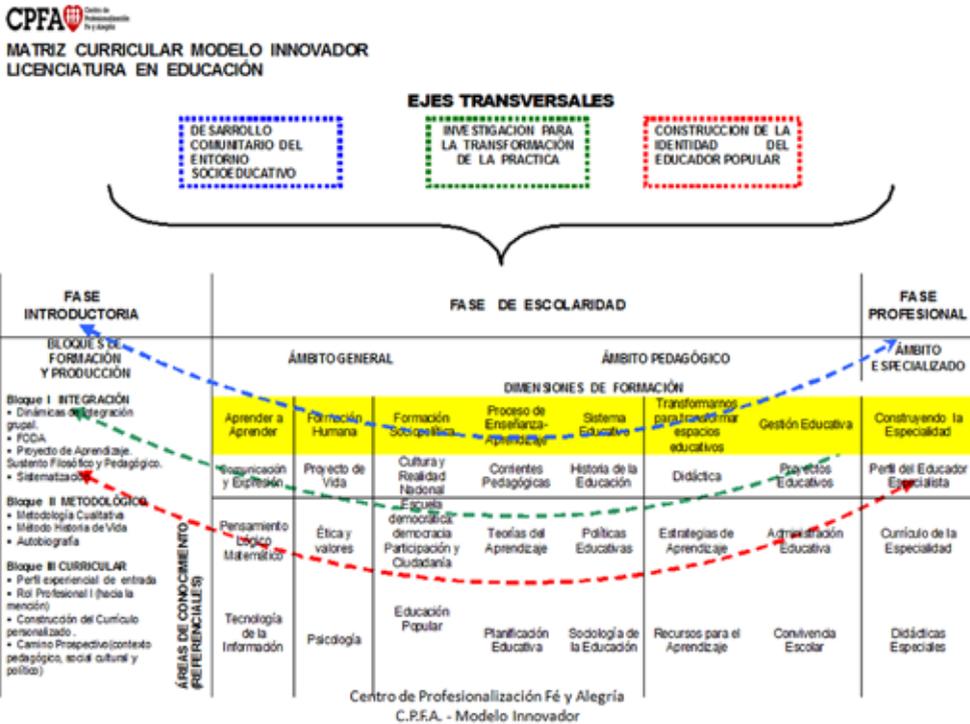


Fig. 3.

Es muy importante resaltar en este proyecto su alcances, sus posibilidades de impactar en forma real la realidad socio educativa de una población vulnerable o marginada desde una apuesta integral y responsable como es la formación profesional de quienes han hecho esta labor durante años partiendo algunos de su formación secundaria con énfasis en pedagogía o bachiller “normalista” como se llama en Colombia; otros con formación técnica pero sin formación pedagógica. Pero un aspecto importante a destacar es que se trata de personas que están unidas a estos proyectos por vocación de servicio, por afinidad en el proyecto de educación popular y que comparten un mismo proyecto de humanidad o de espiritualidad, por tanto corresponde a la dirección de la organización de escuelas de Fe y Alegría potencializar estas capacidades y brindar una formación que sea integral y que vaya hacia los dos niveles el de los niños y niñas y el de la promoción profesional de los docentes y la comunidad en general. Se trata entonces de un Sistema Complejo de Formación en el que se integran la comunidad de aprendizaje –virtualidad–, el diálogo de saberes, la construcción de conocimiento, espiritualidad, sistematización y educación popular (Fig. 4).

Es un modelo innovador para la formación de licenciados en educación basado en la pedagogía de la educación popular, la sistematización de la práctica y la incorporación de nuevas tecnologías a los procesos de aprendizaje y está inspirado en una idea planteada por el fundador de Fe y Alegría que dice: “La formación de educadores en ejercicio es uno de los capítulos de más relieve

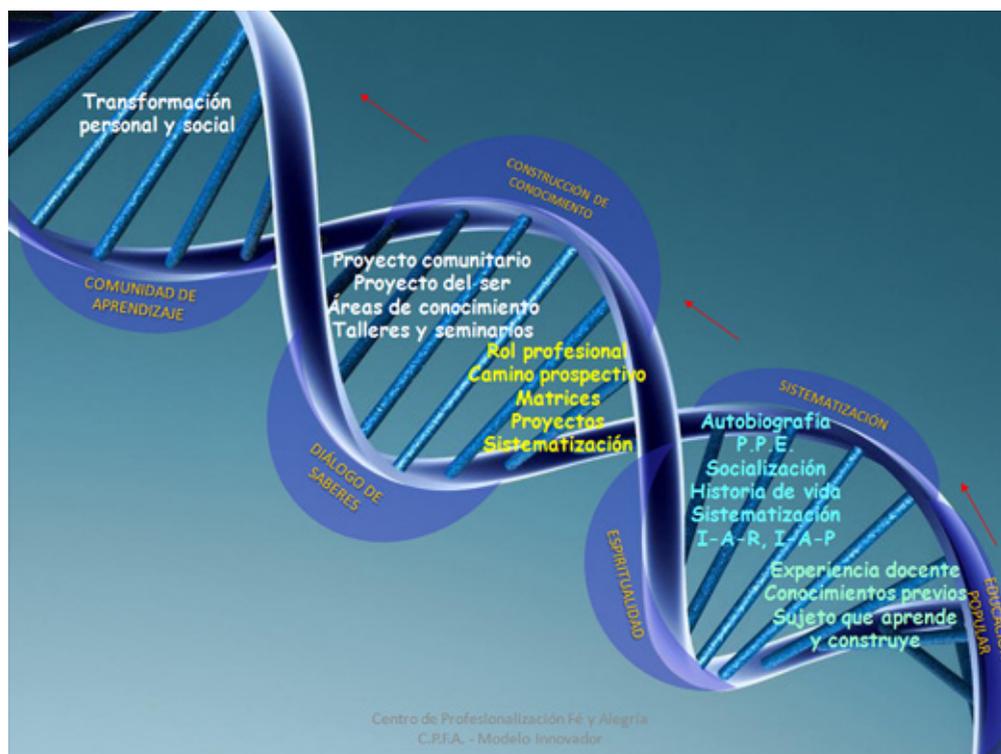


Fig. 4.

en la misión de Fe y Alegría... son los docentes los que llegan a los niños y jóvenes más necesitados... Se vive y se trabaja en formación para la vida” José María Vélaz.

CONCLUSIÓN

Las realidades de América Latina nos obligan a ser creativos y no quedarnos en los discursos dilatorios del deber ser y de pedir o exigir que el sistema político responda a sus obligaciones, sino que es necesario responder de manera concreta con las herramientas disponibles como la virtualidad para mejorar nuestros entornos y empoderar a las comunidades y en este sentido estos proyectos, entre otros, han sido importantísimos en la contribución de cierre de la brecha entre ricos y pobres a partir de la educación.

La experiencia como tutor ha permitido tener muchos aprendizajes, pero lo más importante ha sido el organizar las actividades para lograr que los participantes, profesores de mucha experiencia, puedan compartir expectativas, conocimientos y logren plantear un proyecto en forma colectiva. Cada cohorte es diferente y por tanto además de las habilidades pedagógicas adquiridas en la tutoría en red no están concluidas, sino que hay que seguir aprendiendo. Pero además cuando surgen propuestas como éstas que he destacado brevemente que hacen una apuesta real por la formación que a su vez se ve replicar sobre

la gente más necesitada generando esperanza y renovación de un compromiso especial que traspasa los límites del tiempo y el reconocimiento desde el lugar de la vocación como maestros y acompañantes ignacianos para seguir generando esperanza en la construcción de humanidad y justicia.

BIBLIOGRAFÍA

- Coll, C.; Mauri, T. y Onrubia, J., "Análisis y resolución de casos-problema mediante el aprendizaje colaborativo", *Revista de Universidad y Sociedad del Conocimiento*, 3 (2) (2006), pp. 29-41.
- De Wever, B.; Van Keer, H.; Schellens, T. y Valcke, M., "Roles as a structuring tool in online discussion groups: The differential impact of different roles on social knowledge construction", *Computers in Human Behavior*, 26 (2010), pp. 516-523.
- Hou, T. y Wu, S., "Analysing the social knowledge construction behavioural patterns of an online synchronous collaborative discussion instructional activity using an instant messaging tool: A case study", *Computers and Education*, 57 (2011), pp. 1459-1468.
- Interactive Educational Multimedia (IEM)*, Engel, A. y Onrubia, J. (autores), "Scripting computer-supported collaboration by university students", Barcelona, 2008.
- Karagiorgi, Y. y Symeou, L., "Translating Constructivism into Instructional Design: Potential and Limitations", *Educational Technology & Society*, 8 (2005), pp. 17-27.
- Lechner, Norber, *Los desafíos políticos del cambio cultural*, 2000.
- Levy, Pierre, "Cibercultura y Educación", *Dominical El Colombiano*, 15 de marzo de 1998.
- Liaw, S.S., "Considerations for developing constructivist web-based learning", *International Journal of Instructional Media*, 30 (2004), pp. 309-321.
- Marín, M. y Mateos, A., "The role of new technologies in competence teaching in higher education: the case of accounting", *US-China Education Review*, 7 (3) (2010), pp. 53-61.
- Martí, E., *Representar el Mundo Externamente*, Madrid, Aprendizaje, 2003.
- Montes, J. y Ochoa, S., "Apropiación de las tecnologías de la información y comunicación en cursos universitarios", *Acta Colombiana de Psicología*, 9 (2) (2006), pp. 87-100.
- Onrubia, J. y Engel, A., "Strategies for collaborative writing and phases of knowledge construction in CSCL environments", *Computers and Education*, 53 (2009), pp. 1256-1265.
- Sánz, Rosa María, *Aprender con tecnología. Investigación internacional sobre modelos educativos de futuro*, España, Fundación Telefónica-Ariel, 2012.
- The Quarterly Review of Distance Education*, Du, J. y Xu, J. (autores), "The quality of online discussion reported by graduate students", Miami Beach, Fischler school of education and human service nova southeastern university, 2007.
- UNESCO, *Hacia las sociedades del conocimiento*, UNESCO, 2005.
- <http://estadisticas.cepal.org/cepalstat/WEB_CEPALSTAT/Library/CEPALSTAT/Infograficos_pdf/Inf_1_es.pdf> [consultado 05/05/2013].



RESUMEN

Este artículo presenta el Diplomado en Diseño de Proyectos Virtuales con énfasis en Proyectos de Intervención Social o Proyectos de Aula (DATIC), dirigido a docentes de las universidades regentadas por la Compañía de Jesús en América Latina, AUSJAL, y de otras obras similares, que comparten el mismo proyecto educativo y espiritualidad ignaciana. Es una propuesta de contribución en la formación pedagógica, en la investigación y en el servicio que busca aprender a interactuar con los demás docentes mediante el uso de nuevas tecnologías que permitan diseñar proyectos en forma colaborativa que respondan a las necesidades de la sociedad menos favorecida.

Se presenta la caracterización de dos proyectos que son fruto del trabajo colaborativo del Diplomado, estos son: el "Diplomado virtual en Gerencia Social" y el de "Profesionalización de docentes en ejercicio, a través de comunidades de aprendizaje B-Learning". En ambos se destaca el enfoque del aprendizaje a través de la colaboración y del uso de la tecnología como soporte de formación en función del bien común, permitiendo algunos cambios en la perspectiva del aprendizaje y de la realidad concreta de las personas, el primero en una fase inicial haciendo énfasis en la formación de los gerentes sociales a través del diseño instruccional del curso y el segundo en su fase ya de concreción en relación con la profesionalización docente

Palabras clave: AUSJAL, DATIC, Proyectos Sociales-Aula, Aprendizaje colaborativo, Pedagogía ignaciana.

ABSTRACT

This article presents the certification program "*Design of Virtual Projects with Emphasis on Social Intervention or Classroom Projects (DATIC)*". It is addressed to teachers of universities led by the Society of Jesus in Latin America (AUSJAL) as well as other similar Jesuit institutions who have the same educational project and Ignatian spirituality. This certification program is a contribution to teachers' pedagogical, investigative and service education. Its aim is for them to learn how to interact with other teachers through the use of new technologies to collaboratively design projects that respond to the needs of disadvantaged member of society.

Here we report the characterization of two projects that are the result of collaborative work made in the certification program, whose titles are the "Virtual Certification Program in Social Management" and "Professionalization of teachers through B-Learning". In both we find remarkable learning through collaboration and the use of technology. They support educating for a common good, allowing changes in the learning perspective, and contributing to the reality of people's lives. The former project is an initial stage of education that emphasizes the training of social managers through the instructional design of the course, while the second one is the final stage that emphasizes teachers' professionalization.

Keywords: AUSJAL, DATIC, Social Projects, Classroom Projects, Collaborative Learning, Ignatian Pedagogy.

El buscador de GICES XIX, herramienta digital sobre el cuento español del siglo XIX

MONTSERRAT AMORES
Universitat Autònoma de Barcelona
Montserrat.Amores@uab.cat

El cuento literario ha sido uno de los géneros más cultivados desde el siglo XIX y la prensa periódica fue uno de sus principales medios de difusión. Su progreso durante ese siglo fomentó el auge de la literatura de la época, en especial del cuento, dada su brevedad y fácil acomodo en las páginas de diarios y revistas. A pesar de la abundante producción cuentística, la bibliografía científica ha abordado escasamente el estudio de este género que es una pieza clave del sistema literario de la época. No contamos todavía con el análisis de las narraciones publicadas en las revistas literarias más representativas del siglo XIX. Este trabajo, que sólo puede ser abordado por un equipo de investigadores, facilitaría estudios dedicados a escritores menores, subgéneros, o aproximaciones temáticas cultivadas en la época.

GICES XIX (Grupo de Investigación del Cuento Español del Siglo XIX) lleva a cabo la catalogación y estudio de los cuentos publicados en el *Semanario Pintoresco Español* (1836-1857), *El Museo Universal* (1857-1869), *Museo de las Familias* (1843-1871), *El Laberinto* (1843-1845), *El Siglo Pintoresco* (1845-1848) y *El Mundo Pintoresco* (1858-1860), gracias a los proyectos de investigación HUM2007-63108 y FFI2011-24314. El espacio cronológico comprendido entre 1836 (primer año de publicación del *Semanario Pintoresco Español*) y 1871 (último de *Museo de las Familias*) abarca todo el reinado de Isabel II y una mínima parte del Sexenio Revolucionario, período en el que también predomina el llamado "canon isabelino". Desde el punto de vista literario, se trata, por una parte, de los años de consolidación y desarrollo del romanticismo en España –teniendo en cuenta la evolución que se produjo desde el romanticismo histórico y el impacto de las nuevas tendencias románticas–, por otra, del paso de los códigos predominantes del romanticismo a las nuevas propuestas del realismo. En las páginas de estas revistas escribieron los más reputados escritores románticos, como Eugenio de Ochoa o Juan Martínez Villergas; aquellos considerados precedentes del realismo, como Fernán Caballero o Antonio de Trueba; y los que escribirán bajo las filas del realismo, como Pedro Antonio de Alarcón. El análisis de las narraciones breves publicadas en estas revistas contribuirá decisivamente a la recuperación de autores olvidados y aportará

MONTSERRAT AMORES

«El buscador de GICES XIX, herramienta digital sobre el cuento español del siglo XIX»
Humanidades Digitales: desafíos, logros y perspectivas de futuro. Janus, Anexo 1 (2014), pp. 79-85

datos decisivos sobre la evolución y el desarrollo del género y sus diferentes temáticas en España, mostrará las preferencias por el cultivo de ciertos géneros y pondrá en evidencia el paso gradual del romanticismo a la estética realista. El resultado final será una contribución decisiva a la historia del cuento español que reflejará no sólo la importancia de este género en el panorama literario español, sino también su evolución y desarrollo.

La firme convicción de los miembros de GICES XIX de que la incorporación de las nuevas tecnologías a los estudios literarios es imprescindible para el cabal desarrollo y evolución de esta disciplina fue el fundamental acicate para la creación de una herramienta digital dirigida a los investigadores de la narrativa breve española del siglo XIX en el ámbito universitario, igualmente útil para docentes y estudiantes de enseñanzas medias. Nuestra principal aspiración es que nuestra página web y su buscador se conviertan en un medio idóneo de transferencia de conocimiento.

El principal estímulo que nos decidió a crear una web en la que se incluyera parte de nuestro trabajo como investigadores llegó de la Hemeroteca Digital de la Biblioteca Nacional de España. La facilidad que suponía poder consultar desde casa los textos que debíamos estudiar originó la idea de crear una web de nuestro equipo de investigación que contuviese una intranet a través de la cual poder catalogar los cuentos de las revistas y las entradas de las biografías de los autores, es decir, crear un portal sobre el cuento español del siglo XIX publicado en la prensa periódica, que pudiese ser consultado por cualquier usuario y que fuese una herramienta digital muy útil para los investigadores en este campo. El principal reto técnico para el webmaster es la gestión, actualización y exposición de la base de datos. La estructura de los datos es relacional por lo que se usa *MySQL* para el almacenaje. La gestión del webmaster se hace principalmente mediante *phpmyadmin* y las tareas automáticas mediante *scripts* en *php* y *bash*.

Las ventajas y beneficios de una web construida partiendo de estas premisas son varios:

1. Facilita el trabajo de todos los miembros del equipo, puesto que se puede realizar desde casa o desde otro lugar su trabajo. Sólo se tiene que estar conectado a Internet para acceder a la web del proyecto y a la intranet y, a continuación, llevar a cabo las tareas de catalogación y redacción de los informes de cuentos y autores.
2. Posibilita el trabajo colaborativo, pues el investigador tiene acceso en la intranet a toda la información aportada por los otros miembros del equipo y puede sugerir cambios o modificar sus propios informes.
3. Se trata de un trabajo en curso que se va completando a la vez que se avanza en la investigación. Los miembros del equipo deciden el momento en el que cada uno de los informes puede ser publicado. Conforme ampliamos el número de revistas, los informes de autores van aumentando y también el número de informes de cuentos. El entorno digital permite actualizar los resultados de nuestra investigación.
4. La información llega a través de Internet a los usuarios que consultan nuestra web y acceden a través de un buscador y un directorio (cuentos, autores, traducciones y adaptaciones, años, revista...) a la información que precisa.

LA INTRANET

Como señalaba anteriormente, era esencial crear una intranet, un ámbito de trabajo del grupo de investigación al que sólo pudieran acceder los miembros del equipo mediante contraseña.

La intranet es un espacio web diseñado para facilitar su trabajo. Para su funcionamiento, el trabajo del servidor de interacción con la base de datos está programado en su totalidad en *php*, mientras que la interfaz trabaja con tecnologías web: *HTML*, *CSS*, *java-script* y *jQuery*. La actualización de la base de datos la hacen los investigadores del proyecto mediante esta intranet.

Contiene diferentes espacios (Gestión de Autores, Gestión de Cuentos, Gestión de Traducciones, Gestión de Artículos, Informes, y Administración y Ayuda). Los tres primeros ofrecen acceso a los modelos de informes de autores, cuentos y traducciones con campos específicos para introducir los datos. Los investigadores firman cada uno de los informes de los que son responsables a través de Gestión de Autores e introducen a través de las fichas de escritores los enlaces entre autor y cuento, o bien entre diferentes autores o diferentes cuentos, con el fin de activar los vínculos y la navegación.

Gestión de Artículos alberga la catalogación de las revistas objeto de nuestro estudio no catalogadas anteriormente por otros estudiosos o investigadores, un punto de partida esencial para iniciar el trabajo de análisis de los cuentos. En la actualidad alberga la catalogación de *El Siglo Pintoresco*, *El Laberinto* y *El Mundo Pintoresco*.

En Informes encuentra el investigador el Informe final, tal y como lo verá el usuario. Es importante señalar aquí que los miembros del equipo pueden ver todos los informes, incluso los que no están activados, con el fin de sugerir cambios o modificar los contenidos de su trabajo. Solamente cuando el responsable de un autor activa el informe del escritor, el usuario puede consultar la información desde el directorio o el buscador de la web de GICES XIX.

Finalmente, Administración y Ayuda es un espacio que permite a la investigadora principal gestionar el trabajo de los miembros del equipo: la aplicación permite dar de alta a un miembro y adjudicar autores, modificar autores o eliminar miembros del equipo. También contiene una página con el trabajo realizado por cada investigador y archivos que contienen información detallada sobre cómo elaborar un informe de autor, de cuento, o navegar por la intranet.

LA WEB DE GICES XIX

La página web de GICES XIX se encuentra alojada en un servidor Apache de la Universitat Autònoma de Barcelona y es, asimismo, un subdominio del portal de la Universidad: <gicesxix.uab.es>. Consta de una página principal que contiene diferentes pestañas que dan acceso a la Presentación del grupo de investigación, a las Actividades del Equipo, los Proyectos de los que ha disfrutado y disfruta, se presenta a sus Miembros y sus Publicaciones en los últimos años.

La exposición de la base de datos se hace desde el buscador del proyecto, y permite al usuario consultar la base de datos, que está programada con las mismas tecnologías que la intranet: *php* para el servidor y tecnologías web para el cliente.

El diseño de la web es primordialmente intuitivo, pues el usuario puede navegar sin tener que examinar detenidamente las pantallas, ni especular sobre lo

que podría o no hacer. La página de inicio contiene la que consideramos información básica: Qué es el buscador, qué información contiene y cómo buscar esa información. El catálogo vertebra esta información en tres modelos de informe: cuento, traducción y autor correspondiente a la Gestión de cuento, Gestión de Traducción y Gestión de Autor de nuestra intranet.

Así, el usuario encontrará en ella la breve definición de "cuento" de la que partimos para la catalogación y de las delimitaciones que proponemos, teniendo en cuenta el contexto en el que se publican los textos catalogados. Además, tiene acceso a una breve descripción del contenido de los informes: título y variantes de título; datos completos de la revista en que se publica; el volumen o volúmenes en que aparece posteriormente la narración; si existen o no variantes entre las diferentes versiones del cuento; un resumen del relato; un campo que contiene los temas, motivos y tipos que desarrolla el texto; una breve descripción de los aspectos formales del cuento, la sección de la revista en la que se publica el relato y una clasificación genérica.

En el caso de las traducciones y adaptaciones de narraciones de autores extranjeros vertidos en las revistas estudiadas, también consideramos las versiones prosificadas de textos publicados originariamente en verso, o fragmentos de obras más extensas que, leídos en su versión traducida o adaptada, puedan ser considerados cuentos. El diseño de este informe es similar al de los cuentos, si bien dispone de algunos apartados específicos. En primer lugar, se proporciona, siempre y cuando haya sido posible su identificación, el Título original y el Autor del cuento. En el campo Firmado se indica quién puso su rúbrica al texto; debe tenerse en cuenta que los autores españoles firmaban en ocasiones los cuentos sin especificar su papel de adaptadores o traductores. Por último, en el apartado de Observaciones se ofrecen los datos completos del original, así como otros que el investigador considere pertinentes. Dada la dificultad de identificar el texto original del que procede la traducción, en ocasiones algunos cuentos aparecen bajo el membrete Anónimo, a la espera de que futuras investigaciones identifiquen al autor del texto.

Finalmente, en Autores se ofrece una sucinta información sobre la vida y obra del autor, junto a una breve nota descriptiva de todas sus colaboraciones en la revista. Se facilita también el seudónimo o el título nobiliario por el que era conocido. Los cuentos publicados sin firma se asocian a la entrada de Anónimo. En el caso de los textos firmados con siglas, en el informe aparece encabezado por dichas siglas.

El grado de información que tenemos de cada uno de los autores (algunos desconocidos, otros ampliamente estudiados) es muy diverso, de ahí la heterogeneidad de estas biografías, semblanzas concebidas como un complemento de los cuentos documentados. La biografía contiene, además, un párrafo dedicado a sus colaboraciones en la revista. Seguidamente se ofrece un listado de dichas colaboraciones dividido en dos apartados: Cuentos (es posible acceder a cada uno de los informes de las narraciones mediante los hipervínculos) y Otras colaboraciones. El informe se cierra con una Bibliografía donde únicamente se señalan las referencias bibliográficas utilizadas para la confección del informe. Asimismo, debe tenerse en cuenta el carácter provisional de estas fichas, que irán completándose a medida que se incorporen colaboraciones de otras revistas. No se refiere la biografía de los autores extranjeros cuyos cuentos se tradujeron en la publicación. Todos estos informes pueden imprimirse.

EL DIRECTORIO Y EL BUSCADOR

El usuario puede acceder a los datos mediante nuestro buscador que contiene un Directorio y un motor de búsqueda por palabras.

El Directorio permite buscar los informes por cuentos (españoles y traducidos) ordenados alfabéticamente, por autores (se señalan los seudónimos y también si se trata de un autor extranjero), por género, por año o por revista. La consulta de cada uno de los informes le permite navegar mediante los enlaces y consultar los informes de cuentos o de otros autores. La búsqueda por años permite ver los informes de todos los cuentos publicados en un mismo año, mientras que en la de Género el usuario puede marcar una o más clasificaciones genéricas para conseguir una lista de cuentos consignados dentro de su selección. Finalmente, la consulta por revistas permite seleccionar el título de una revista y ver los informes de cuentos publicados en dicha publicación.

Por su parte, la Búsqueda simple proporciona una detallada información sobre los cuentos aparecidos en las revistas españolas del siglo XIX objeto de nuestro estudio, que se amplía y actualiza progresivamente conforme el equipo de trabajo completa el estudio de cuentos y de autores. El sistema, mediante un diseño intuitivo, ofrece los resultados de la búsqueda de una palabra o de una palabra con un comodín (*), combinando el sistema de listados, en este caso tres: cuentos, autores y años. El visitante puede a continuación decidir por cuál de los listados ordenar los resultados y consultar directamente, el informe de los cuentos resultado de la búsqueda, de los autores o de los años consignados.

Como señalaba anteriormente, la web de GICES XIX y su buscador se presentan, pues, como una potente herramienta de trabajo que ofrece nuestro Grupo de investigación a otros investigadores o usuarios. Gracias a ella el visitante tiene acceso a una serie de conocimientos que pueden ser punto de partida de futuros estudios. A continuación proponemos algunos de los objetivos finales de nuestro proyecto.

1. Ofrece el índice de autores españoles que publicaron cuentos en la prensa periódica y de sus principales aportaciones al género y al periodismo, gracias al cual se puede sistematizar una panorámica histórica del cultivo del cuento literario en el período establecido. De esta manera se estará en condiciones de realizar trabajos monográficos sobre los autores que publican en las páginas de las revistas objeto de catalogación y estudio.
2. El visitante puede encontrar una cronología de los cuentos que permite establecer, a lo largo de los años en que se centra el proyecto, la evolución estética y temática que pudo llegar a experimentar cada uno de los autores escogidos, así como los cambios sustanciales que sufrieron las revistas en su seno, de acuerdo con las premisas literarias que fueron prevaleciendo en la etapa histórica seleccionada.
3. La información que proporciona el Buscador facilita el estudio temático de los relatos, comprobando la recurrencia de temas, de tipos de personaje y de motivos literarios en las revistas estudiadas, de acuerdo con los géneros predominantes en dicho período histórico, como el cuento fantástico, legendario, histórico o costumbrista. Tal análisis facilitará un estudio a nivel diacrónico y sincrónico, tanto entre los diversos escritores seleccionados, como entre las publicaciones.

4. Gracias al análisis de las características formales y estéticas de los relatos y al período cronológico seleccionado para el estudio, entre 1836 y 1870 aproximadamente, la época de mayor consolidación del período romántico y, por otro, el inicio de la estética realista, resulta de gran interés establecer, además, las características que se mantuvieron y las innovaciones formales que aparecieron entre un movimiento literario y otro.
5. La información suministrada en la web es una base sustancial para el examen de la orientación ideológica de las revistas dependiendo del contexto histórico en el que se desarrollan, así como el estudio del papel preponderante que tuvieron los sucesivos directores. Como ya se ha comentado, el período seleccionado abarca todo el reinado de Isabel II, precisamente caracterizado por la heterogeneidad en las directrices ideológicas de su mandato. Por todo ello, resulta sumamente interesante contrastar las etapas moderadas (1843-1854), las más liberales (1854-1856) o, incluso, la misma decadencia de la monarquía (1866-1868) con las diferentes vías de pensamiento que fueron predominando a lo largo de estos años en las páginas de las revistas estudiadas.

Estos, como señalaba, son sólo algunos de los posibles ámbitos de estudio para los que nuestro buscador es una herramienta digital idónea. A pesar de las dificultades y las contrariedades que surgen a cada paso para obtener nuestros objetivos, los miembros de GICES XIX, con su técnico informático, estamos convencidos de que este, el de Internet, es el medio idóneo para conseguir la mejor visibilidad y difusión de nuestra investigación.



RESUMEN

El objetivo principal de la siguiente comunicación es la presentación de la web de GICES XIX (Grupo de Investigación del Cuento Español del Siglo XIX) y muy especialmente de su buscador, herramienta digital dirigida a los investigadores de la narrativa breve española del siglo XIX y al público universitario (<http://dfe.uab.es/gicesxix/>). Tras indicar los objetivos del actual equipo de investigación, se presenta la intranet de GICES XIX (el ámbito interno de trabajo del grupo de investigación) y su organización: la administración de los usuarios, la pantalla de trabajo, el diseño de los informes de autores, cuentos y traducciones, y los procedimientos de introducción de datos. Señala las ventajas de la intranet y su modo de gestión, pues permite llevar a cabo las tareas de catalogación y redacción de los informes de autores y de los cuentos que se publicaron en las páginas de las revistas en las que trabajan los miembros del equipo.

Seguidamente se describe la interfaz del buscador para el usuario: los diferentes modos de navegación que ofrece y los distintos sistemas de búsqueda de datos, gracias a su intuitivo diseño. Ofrece la posibilidad de acceder a toda la información referida sobre un cuento, un autor, un subgénero de cuento, los cuentos publicados en un año concreto... y navegar a través de los enlaces. Todo ello facilita el estudio del género desde numerosos y diferentes enfoques.

Palabras clave: Cuento, siglo XIX, prensa, catálogo, base de datos.

ABSTRACT

The main purpose of the following paper is to introduce the website of GICES XIX (19th-century Spanish Short Stories Research Group) and, especially, its web search en-

gine, which is a digital tool aimed at researchers of 19th-century Spanish Short Narrative and university students (<http://dfe.uab.es/gicesxix/>). After stating the objectives of the current research team, the paper shows the intranet of GICES XIX (the research group's internal work area) and its organization: the administration of users, the work screen, the design for the reports of authors, stories and translations, and data entry procedures. It points out the advantages of the intranet and its management mode, as it allows the group members to carry out the tasks of cataloging and writing the reports of authors and stories published in the magazines on which they work.

After that, the paper describes the user interface of the search engine: the different navigation modes it offers, thanks to its intuitive design. It also provides the possibility of accessing all the information related to a story, an author, a subgenre of story, the short stories published in a particular year and... navigating through the hyperlinks. All these functions facilitate the study of this specific literary genre from a variety of approaches.

Keywords: Short story, Tale, 19th Century, Press, Catalogue, Database.

Extracción y organización del conocimiento de etiquetados. Aplicación a etiquetados en repositorios digitales sobre arte¹

GONZALO A. ARANDA CORRAL

Departamento de Tecnologías de la Información, Universidad de Huelva
garanda@us.es

JOAQUÍN BORREGO DÍAZ

Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Sevilla
jborrego@us.es

JUAN GALÁN PÁEZ

Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Sevilla
juangalan@us.es

INTRODUCCIÓN

Las herramientas y servicios que la Web Social ha popularizado ofrecen a la comunidad académica tecnologías para avanzar en el análisis de los recursos digitales de investigación. Una de las primeras etapas en esa adopción consiste ofrecer repositorios digitales para que cualquier investigador tenga fácil acceso a los recursos. Los recursos se disponen bajo indexación, y suelen clasificarse de acuerdo a categorías. Junto con la categorización, una de las técnicas más populares y útiles para su clasificación y exploración es el etiquetado de dichos recursos.

El etiquetado permite el uso de *folksonomías* (vocabularios con cierto consenso en la comunidad que marca un tipo de recursos). La folksonomía, que se puede entender como un tipo de ontologías, permite al usuario navegar u organizar el conocimiento que el repositorio gestiona.

Sin embargo, este tipo de *ontologías emergentes* sufren de ciertas limitaciones y deficiencias. Por un lado, al ser un vocabulario con una intencionalidad dependiente del usuario que etiqueta, es habitual que exista cierta *heterogeneidad semántica*: ciertas etiquetas representan conceptos distintos para distintos etiquetadores. La heterogeneidad semántica es uno de los problemas intrínsecos al uso de etiquetados, y que impide que un usuario se aproveche con fiabilidad del etiquetado de otro. Por otro lado, junto a etiquetas con utilidad para la comunidad, nos encontraremos con etiquetas que son usadas raramente, que no

¹ Financiado por el Proyecto de excelencia TIC-6064 de la Junta de Andalucía *Conceptos emergentes en sistemas complejos. Aplicaciones en entornos urbanos y en complejidad cultural (eComplexity)*, cofinanciado con fondos FEDER.

guardan relación semántica con el objeto etiquetado (por ejemplo *to read* para marcar una tarea futura) o cuyo significado real no es conocido por gran parte de la comunidad. A pesar de las limitaciones, los etiquetados compartidos es una solución potencial para acercar a la comunidad el conocimiento semánticamente organizado.

El problema anteriormente descrito es muy similar al que se enfrenta un usuario cuando intenta recabar información de un repositorio etiquetado. Cuando recurrimos a repositorios ajenos los etiquetados nos ofrece una rápida navegación. Debemos "entender" el significado de las etiquetas. En este caso, debemos *reconciliar* nuestro conocimiento con el conocimiento del repositorio. Es decir, establecer relaciones semánticas entre nuestros conceptos y nuestra manera de etiquetarlos y los del repositorio. Para que esta tarea (eminentemente racional y que realizamos de manera inconsciente) tenga éxito el etiquetado que exploremos debe cumplir ciertos criterios de consistencia y robustez.

Objetivo del trabajo

En este trabajo nos planteamos el uso del Análisis Formal de Conceptos (FCA) para estudiar la adecuación de los etiquetados en repositorios digitales. Propondremos una medida para su adecuación, basada en el estudio de otros objetos similares (redes semánticas, lenguajes de descripción, etc.) y aplicamos el estudio a algunos repositorios culturales y de arte.

ETIQUETADOS

El éxito del etiquetado en la Web Social se debe a que no tiene, en principio, ningún tipo de limitación de uso personal. Etiquetar es una tarea que en diferentes servicios o portales se consideran de distinta forma. Fundamentalmente se pueden usar de cuatro formas (Smith, 2007): Para manejar información personal (navegar entre los recursos que nos parecen más interesantes y que hemos etiquetado), como método social para marcar objetos digitales y así facilitar su compartición, o incluso para mejorar la experiencia de usuario en el comercio electrónico. Así, el usuario puede explorar su *personomía* así como las *personomías* de otros usuarios: un usuario puede ver los recursos que otros usuarios han subido y/o etiquetado (Jäschke, 2008).

El etiquetado como representación semántica

Sin embargo, y debido a las diferentes formas de usar las etiquetas, el manejo global, fiable y automatizado del conjunto de todas las etiquetas utilizadas en una comunidad (la *folksnomía*) es un desafío. Cuando el usuario añade una etiqueta a un recurso es difícil atrapar todos los matices que el usuario pretendía representar. Como se argumenta en (Golder *et al.*, 2006), el etiquetado es una tarea para dotar de sentido los recursos que poseemos o visitamos, y pretende categorizar de manera que existe un significado emergente que se produce con esa actividad (Weick *et al.*, 2005). Como conocimiento emergente, la frontera de esos conceptos que manejamos cuando elegimos etiquetas son vagos, lo que puede producir que algunas etiquetas sean de dudosa utilidad o adecuación. Como consecuencia de eso, un etiquetado personal puede ser poco útil como bien público (Golder *et al.*, 2006).

Aunque a distinta escala, la etiquetación de objetos por parte de una comunidad puede presentar el mismo problema. Sin embargo, en este último caso la dificultad es usualmente resuelta por la Inteligencia Colectiva: cuando la comunidad etiqueta globalmente se tiende a unificar el uso de la etiqueta, y el conjunto de etiquetas más comunes asociadas a un objeto o conjunto de objetos representa una descripción de un cierto concepto (Halpin *et al.*, 2007). De hecho, estas etiquetaciones globales facilitan la construcción de sistemas de recomendación (Carmel *et al.*, 2010).

Las humanidades digitales pueden y deben explotar esas facilidades para, entre otras tareas, difundir repositorios digitales para que los investigadores tengan acceso a la información. El etiquetado colaborativo, como forma de indexación social, es un método natural y directo para que laboratorios y grupos de investigación permitan la navegación entre sus recursos de manera independiente a la pura clasificación o taxonomía. El conjunto de etiquetas utilizadas, en el caso de temas específicos, es una clasificación social atendiendo a criterios concretos y muy relacionados con el tema al que el repositorio está dedicado, usualmente construida por especialistas. Es decir, la inteligencia colectiva ayuda a superar las limitaciones indicadas en la sección anterior.

Desde este punto de vista, la etiquetación colaborativa ofrece una alternativa pragmática al uso de ontologías (para la Web Semántica). Las ontologías en Tecnologías de la Información son representaciones formales de la conceptualización del dominio de discurso, y su uso (al menos en las primeras etapas de digitalización y gestión de recursos culturales) presenta varias dificultades: se necesita conocer los rudimentos de ese tipo de teorías, se presupone que los usuarios comparten esa ontología (es decir, participan del consenso con el que fue construida), etcétera.

Es indudable, no obstante, que la adopción de tecnologías de la Web Semántica para organizar el conocimiento de repositorios digitales es un paso fundamental para la interoperabilidad de los repositorios y la organización del conocimiento que contiene. Fundamentalmente, resuelve el problema de la heterogeneidad semántica y provee definiciones compartidas de los conceptos. Esa fiabilidad contrasta con las organizaciones personales de la información (por ejemplo, los etiquetados de los usuarios de *Delicious* (<http://delicious.com/>), que necesitan reconciliar su conocimiento con el de otros usuarios para aprovechar su información. En (Aranda Corral *et al.*, 2010) proponemos un método para resolver esta dificultad, basado en agentes y usando técnicas de Análisis Formal de Conceptos. Finalmente, es necesario destacar que sí existen proyectos para estandarizar la semántica de etiquetas, como por ejemplo *Faviki* (<http://www.faviki.com>) y *CommonTag* (<http://commontag.org>).

Heterogeneidad semántica dependiente del contexto

Analizando las limitaciones del etiquetado colaborativo, existen dos problemas fundamentales. El primero consiste en la dependencia de la etiqueta del contexto en la que se usa. Si en el repositorio se manejan objetos muy diversos, es posible que usuarios distintos usen la misma etiqueta para referirse a conceptos distintos. Este problema limita tanto la efectividad del etiquetado como su adecuación. Es la denominada *Heterogeneidad Dependiente del Contexto* (HDC). La segunda limitación es la que denominamos *Ambigüedad Clásica* (AC), inherente

del lenguaje y de los distintos niveles de abstracción/generalidad que distintos usuarios consideran al etiquetar (Golder *et al.*, 2006; Tanaka *et al.*, 1991).

La segunda, AC, no es una limitación crítica si los usuarios etiquetan, por ejemplo, urls (puesto que el contenido del sitio web permite la desambiguación). Utilizando técnicas de análisis clúster se pueden localizar las distintas acepciones de la etiqueta ambigua (Au Yeung *et al.*, 2009). Sin embargo, la HDC está asociada a la *estructura conceptual* que el usuario utiliza pero no está explícita en ninguna parte del sistema.

Como el HDC obstruye cualquier aproximación al consenso sobre las etiquetas, debemos diseñar algún método para intentar solventarla, usando herramientas inteligentes para recuperar el conocimiento y describirlo (Jung, 2009). Llegados a este punto, existen dos perspectivas para atacar el problema. La primera trata de interoperabilidad semántica *global* entre toda la comunidad, y la segunda es facilitar puentes semánticos para que un usuario pueda aprovechar de manera fiable el trabajo de otro, por ejemplo permitiendo que entienda el etiquetado de otro. La primera requiere un tratamiento global y un análisis de la folksonomía, que es la adoptada en este artículo: como paso previo se debería estudiar la calidad del etiquetado proporcionado por la comunidad para el repositorio concreto. La segunda perspectiva se centra en encontrar la conceptualización oculta en la actividad de etiquetado de cada usuario para establecer *traducciones automáticas* de los etiquetados.

Existen proyectos para unificar los etiquetados (y convertirlos de ese modo en metadatos), como por ejemplo DTSIL (Khou *et al.*, 2012) cuyo objetivo es integrar diversos repositorios en una infraestructura de información única, accesible y que admita interfaces usables, utilizando etiquetas que son elegidas mediante extracción de información de los metadatos. Otro proyecto relacionado es *Rich Tags* (<http://research.mspace.fm/projects/richtags>) que se aprovecha del contexto (la comunidad que mantiene el repositorio) para facilitar la navegación/exploración usando etiquetas que, debido a la especificidad del repositorio, no tienen problemas de heterogeneidad semántica, es decir, son etiquetas *semánticas* en ese sentido.

ANÁLISIS FORMAL DE CONCEPTOS

La convergencia de la Web Social hacia una Web de conocimiento depende de cómo tratar con técnicas semánticas formas de organización del conocimiento como el etiquetado. Una herramienta muy útil para dirigir esa convergencia es el Análisis Formal de Conceptos (FCA). Según R. Wille (Ganter *et al.*, 1997), FCA matematiza el concepto filosófico de concepto como una unidad de pensamiento compuesto de dos partes: la extensión y la intención. La extensión está compuesta por todos los objetos que pertenecen al concepto, mientras que la intención comprende todos los atributos comunes a los objetos de la extensión. Partiendo de esta idea, FCA se desarrolla como una rama de la matemática aplicada cuyo objetivo es descubrimiento, extracción y organización de conocimiento a partir de datos cualitativos mediante redes (jerarquías) de conceptos. La teoría provee técnicas para poder razonar con el conocimiento extraído. Es decir, FCA es una aproximación a la minería de datos con una componente semántica (conceptos formales) y una componente lógico-computacional (reglas de asociación y razonamiento automático). Introduzcamos brevemente los rudimentos de FCA y su utilización sobre etiquetados.

La unidad básica de datos es el denominado *contexto formal*. Un contexto formal $M = (O, A, I)$ está compuesto de dos conjuntos, O (objetos) y A (atributos), junto con una relación entre éstos. Los contextos se pueden representar mediante una tabla de valores 0,1. En la figura 1 (izquierda) se representa un ejemplo sencillo de *universo de peces* junto con tres atributos. Dado un conjunto X de objetos (o Y un conjunto de atributos) se define la *intención* de X (respectivamente la *extensión* de Y) como $X' = \{ a \in A \mid oIa \text{ para todo } o \in X \}$ (respectivamente $Y' = \{ o \in O \mid oIa \text{ para todo } a \in Y \}$).

El objetivo principal de FCA es el cálculo de los conceptos extraídos del contexto. Un contexto formal corresponde a la idea anteriormente esbozada: es un par (X, Y) tal que $X' = Y$ e $Y' = X$. Por ejemplo, el *retículo* de todos los conceptos del contexto de peces se muestra en la figura 1 (derecha). Para leer los conceptos de la jerarquía, representados cada uno por un nodo, hay que recorrer el retículo desde el nodo hasta el concepto más general para obtener la intención y hacia el concepto más específico para obtener la extensión. Por ejemplo, el concepto más específico del retículo sobre peces es $(\{eel\}, \{Coast, Sea, River\})$ que correspondería al concepto *pez eurihalino*. Este ejemplo ilustra claramente que el retículo de conceptos contiene conceptos para los que no disponemos de un término que lo denote. Es decir, hemos descubierto un nuevo concepto. El interés de éste debe ser estudiado por el especialista en la materia que describía el contexto.

Para aplicar FCA sobre etiquetados debemos adaptar al formato de contexto formal el entorno formado por etiquetas, usuarios y recursos etiquetados. La aproximación general se puede ver en (Jäschke, 2008), donde utilizan los denominados *conceptos triádicos*. Puesto que en nuestro caso deseamos analizar la estructura general del repositorio etiquetado, no es necesario hacer la distinción del usuario que ha etiquetado el recurso. Siguiendo con el ejemplo, en la figura 2, se describe la estrategia general para aplicar FCA a etiquetados: los ítems etiquetados (en este caso, de *Delicious*) son los objetos y sus etiquetas los atributos asociados. A partir del contexto así construido se calculan los conceptos que se extraen y se obtiene la jerarquía de conceptos, que representa una jerarquía de conceptos sobre el universo de discurso de ese recurso (en nuestro caso, de los peces).

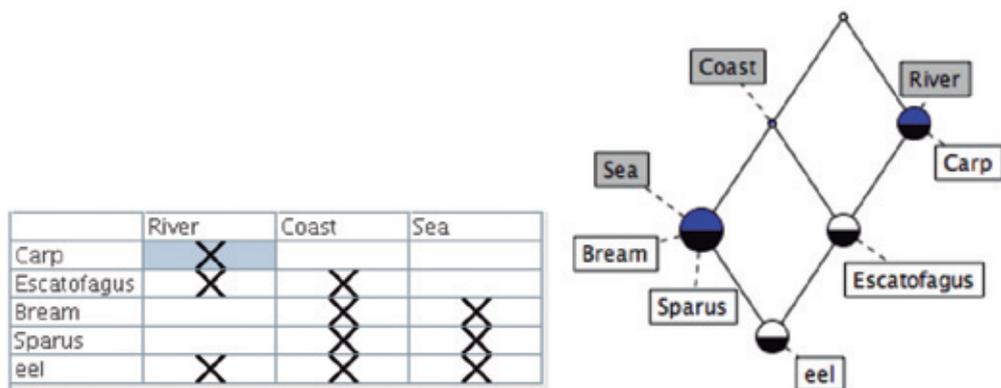


Fig. 1. Contexto formal sobre peces (izquierda) y su retículo de conceptos asociado (derecha).

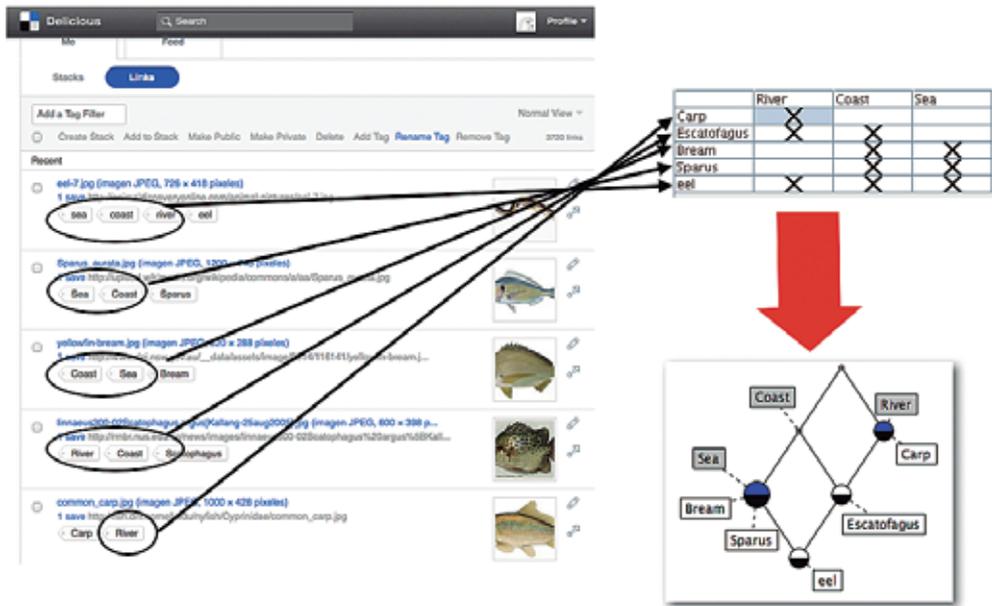


Fig. 2. Extracción del retículo de conceptos a partir del etiquetado de un usuario.

CONCEPTUALIZACIÓN LIBRE DE ESCALA

Llegados a este punto, y antes de analizar repositorios concretos mediante FCA, es necesario encontrar criterios que nos indiquen hasta qué punto el etiquetado es adecuado. Téngase en cuenta que el análisis debe ser independiente del área que abarca el repositorio y del área de conocimiento donde se enmarca el contexto. Recuérdese que FCA produce incluso conceptos que la propia etiquetación no es capaz de asignarle un nombre. Por tanto, cualquier medida de adecuación debe considerar la estructura de los conceptos que abarca una vez aplicado el proceso de conceptualización descrito anteriormente. Para obtener ese test de adecuación debemos analizar la estructura del retículo de conceptos como grafo (red) semántica.

El primer y fundamental análisis se basa en estudiar la distribución de los grados de conexión que posee cada concepto, es decir, con cuántos conceptos está relacionado. Esa distribución es un factor característico de las redes semánticas y ha sido extensamente estudiada para gran variedad de casos. Veamos primero la distribución que caracteriza la complejidad de la red de conceptos. Una vez estudiada, debemos compararla con otras redes semánticas existentes como por ejemplo la asociada a *WordNet*, que representa un caso de éxito, nuestro lenguaje. Parece natural conjeturar que si la estructura de los conceptos que se extrae es similar a la de los casos de éxito, entonces es adecuada (y útil).

Al contrario que otras redes semánticas (Motter *et al.*, 2003), nos enfrentamos a una red con relaciones que no están acotadas por el propio lenguaje, pues FCA extrae todos los conceptos, incluso los que no están definidos en el lenguaje original (puede que ni siquiera dispongamos de una definición explícita). Es tarea del especialista en el tema del repositorio el que debe interpretar los que

considere interesantes. Por tanto, el retículo será una red semántica compleja. De hecho, durante el resto del artículo veremos que en el caso de repositorios con etiquetaciones adecuadas, su topología responde a la de una *red compleja libre de escala*. Una distribución libre de escala es una distribución de los grados de conexión de tipo $P(k) \sim ck^{-\gamma}$, donde $P(k)$ denota la proporción de nodos de la red que tienen k conexiones y c es una constante. Es decir, el número de nodos con alta conectividad es muy pequeño, y existen una gran cantidad de nodos con pocos vecinos. El parámetro γ suele verificar $2 < \gamma < 3$, aunque ocasionalmente pueden encontrarse valores fuera de ese intervalo.

En (Aranda Corral *et al.*, 2012a) estudiamos la siguiente característica, que parafraseamos aquí para analizarlo en el caso de etiquetados, denominada *Hipótesis de la conceptualización libre de escala* (SFCH):

El conjunto de etiquetas asociado al repositorio es adecuado si induce una red libre de escala.

Es decir, si el conjunto de etiquetas induce una estructura de conceptos similar a la de una red semántica de las que emergen en el estudio del lenguaje. Por tanto, podemos considerar FCA como una forma de estudiar el lenguaje emergente de la etiquetación. La hipótesis SFCH se ha testado en casos muy diversos, y se ha comprobado que los retículos provenientes de contextos aleatorios no exhiben esa distribución (Aranda Corral *et al.*, 2012a).

Es necesario en este momento analizar un poco más la existencia de conceptos que no tienen nombre concreto. Es natural pensar que el uso de términos específicos por parte del especialista tiene influencia en las relaciones entre conceptos. Como el retículo de conceptos abarca todo potencial nuevo término del lenguaje (aquel que representaría a un nuevo concepto), la topología de la red será compleja. En (Motter *et al.*, 2003) se estudia la red semántica entre conceptos expresados por términos (en el idioma inglés), mostrando que la topología es libre de escala. En (Motter *et al.*, 2003) se muestra que la conectividad sigue una ley libre de escala $P(k) \sim ck^{-\gamma}$, con $\gamma = 3.5$. En las redes construidas en FCA se tiene que $\gamma > 3.5$ debido a que no está acotada por el propio lenguaje. En (Barabasi *et al.*, 2002) se analiza la red que representa las palabras y la relación de sinonimia, estimando en este caso $\gamma = 2.8$. Otro ejemplo es la red semántica asociada el Tesoro de Roget (Roget, 1911) estudiada en (Steyvers, 2005), para el que $\gamma = 3.19$.

LA ESTRUCTURA CONCEPTUAL DEL LENGUAJE COMPLETO: WORDNET

Para analizar la bondad de un conjunto de etiquetas, es interesante observar cómo se comporta, desde el punto de vista de los conceptos que define, el propio lenguaje que usamos. Con este objetivo usaremos la base de datos léxica *WordNet* (<http://wordnet.princeton.edu/>).

WordNet es una gran base de datos léxica de palabras del Inglés, de uso libre. En esta base de datos, nombres, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos cognitivos (llamados *synsets*), donde cada uno de estos *synsets* expresa un concepto diferente del lenguaje. Es decir, en esta base de datos las palabras se relacionan entre si mediante su significado y por tanto dos palabras estarán relacionadas si comparten una o más acepciones.

Para realizar el análisis de la estructura conceptual, primero construimos el contexto formal asociado a la base de datos. En este paso, consideraremos las

palabras y los *synsets* como objetos y atributos (respectivamente) del contexto. Una vez creado el contexto, calculamos a partir de este, el retículo de conceptos asociado (la estructura conceptual bajo estudio).

Se han considerado tres conjuntos de datos diferentes relacionados con WordNet para ser analizados. Por un lado se ha tomado WordNet al completo, y por otro dos subconjuntos de este, el de los adverbios y el de los verbos (en la tabla 1 se muestran las principales características de los conjuntos de datos usados). De esta forma veremos que las propiedades de todo el conjunto se mantienen en cada una de sus partes.

Tabla 1. Principales parámetros para cada conjunto de datos. Donde |O|, |A| y |I| se refieren al número de objetos, atributos y relaciones (respectivamente) del contexto. *Densidad* indica la densidad del contexto, |CL| el número de conceptos del retículo y <k> el grado medio de los nodos (conceptos) del retículo.

	O	A	I	<i>Densidad</i>	CL	<k>
WordNet completo	155.287	117.659	206.941	0,001%	110.663	2,406
Adverbios sólo	4.481	3.621	5.580	0,03%	3.529	2,187
Verbos sólo	11.529	13.767	25.047	0,02%	12.222	2,967

En la figura 3 se muestra la distribución de grados asociada a los tres conjuntos de datos descritos. Como se puede observar todos ellos cumplen la hipótesis SFCH. Estas estructuras nos servirán para a continuación estudiar los conjuntos de datos relacionados con etiquetados. Además es interesante mencionar que los *conceptos formales* extraídos de WordNet coinciden en gran parte con los conceptos del lenguaje.

ANÁLISIS DE ETIQUETADOS DE REPOSITORIOS DIGITALES DE ARTE E HISTORIA

Para ilustrar el análisis, hemos elegido dos repositorios digitales de humanidades, el de *Baroque Art* (<<http://baroqueart.cultureplex.ca/>>) del laboratorio *CulturePlex* (<<http://www.cultureplex.ca/>>) y el de *Gothic Past* (<<http://www.gothicpast.com/>>). Compararemos su estructura con la conceptualización asociada a subconjuntos notables de WordNet.

El Barroco Hispánico

El proyecto *The Hispanic Baroque: Complexity in the first Atlantic culture* (<<http://www.hispanicbaroque.ca/>>) es un proyecto multidisciplinar desarrollado por un grupo de investigadores de universidades de diferentes países, financiado por el *Social Sciences and Humanities Research Council of Canada*. Con una duración de siete años, el proyecto pretende estudiar el origen, evolución, transmisión y efectividad de los patrones de comportamiento y representación del barroco en el mundo hispánico.

Los objetivos de este proyecto son: identificar y describir los patrones más comunes y resistentes en entornos diversos; establecer su relación con procesos de identidad y organización sociales; analizar las tecnologías de ámbito cultural

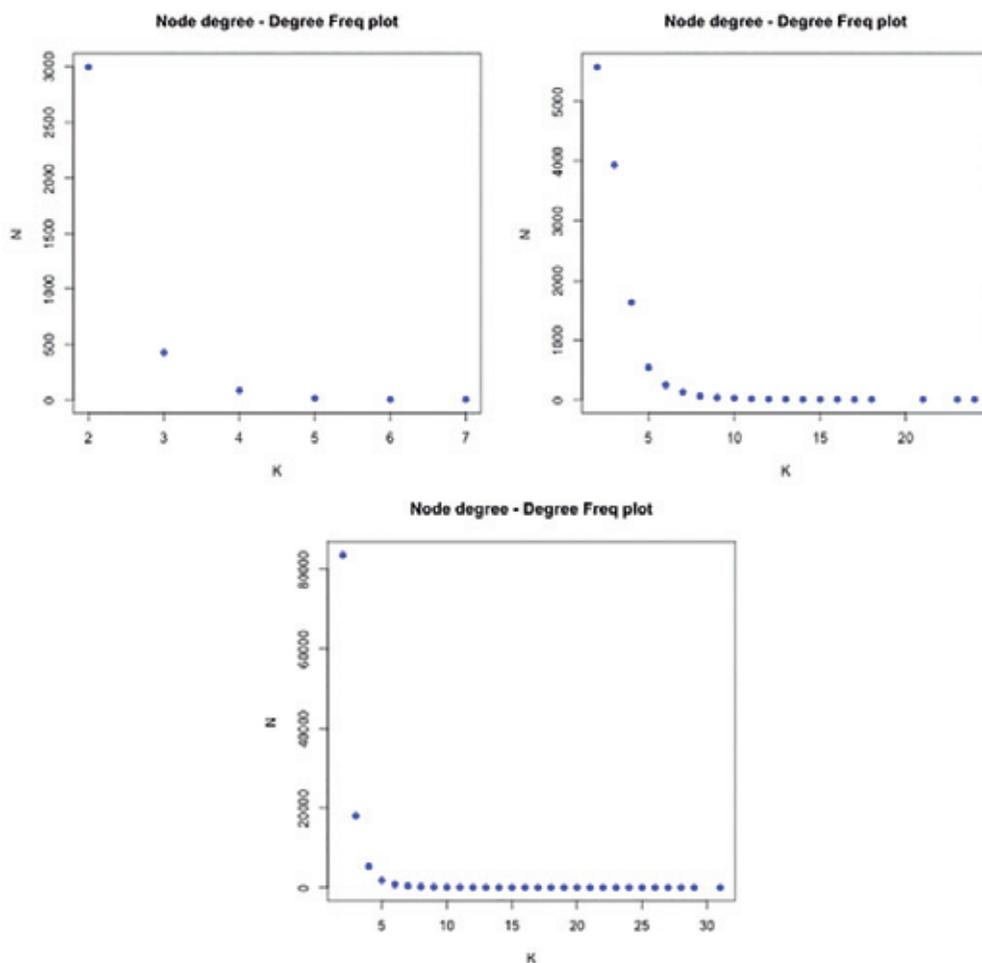


Fig. 3. Distribución de grados de la estructura conceptual asociada al los conjuntos de datos de WordNet: adverbios (arriba a la izquierda), verbos (arriba a la derecha) y WordNet completo (abajo).

que han hecho posible esta adaptación del barroco; determinar su efectividad basada en la reaparición del Neo-barroco en el mundo contemporáneo; desarrollar nuevas herramientas, con la intervención de otras disciplinas, que permitan reforzar los métodos de investigación en humanidades. Este proyecto cubre tres dimensiones fundamentales: investigación, formación de estudiantes y difusión del conocimiento obtenido.

Aunque el proyecto es de gran envergadura, con secciones de diverso ámbito. En este trabajo hemos usado como recurso a analizar la base de datos de obras de arte *Baroque Art*. Esta base de datos proporciona un etiquetado de obras de arte, que es uno de los productos del proyecto *Hispanic Baroque*. De este etiquetado destacamos que ha sido desarrollado por un equipo de personas cerrado y reducido, siguiendo unas directrices comunes preestablecidas y

partiendo de un vocabulario (etiquetas) también cerrado y preestablecido (ontología). El conjunto de datos bajo estudio está formado por aproximadamente 11.000 obras y 200 etiquetas.

En la figura 4 (izquierda) se muestra la distribución de grados de la estructura conceptual asociada al etiquetado de obras del barroco, que también cumple la hipótesis SFCH. Es interesante observar la diferencia que hay en la forma de la distribución, al inicio de esta, con respecto a las de WordNet. Esto se debe a que el lenguaje por sí sólo presenta una estructura muy ordenada, como se comentó en el apartado anterior, mientras que el etiquetado no.

Gothic Past

Gothic Past: Visual Archive of Gothic Architecture and Sculpture in Ireland es un archivo de acceso libre para el estudio de las construcciones medievales de Irlanda. El archivo incluye entre otros elementos, imágenes, etiquetas, descripciones detalladas, etc. Es un proyecto colaborativo entre varias instituciones de Irlanda y financiado por el *Irish Research Council for the Humanities and Social Sciences (IRCHSS)* y el *Irish Heritage Council*. Incluye imágenes y recursos de las principales colecciones de historia, arte y arquitectura de Irlanda. El proyecto está construido sobre la plataforma de software libre *Omeka* (<<http://omeka.org/>>) que facilita la gestión de la información además de su visualización. Además permite de forma sencilla la colaboración con el proyecto de usuarios externos.

Este aspecto colaborativo, por lo que es sumamente inteligente trabajar con su etiquetado de monumentos históricos, ya que nos permitirá compararlo con el caso anterior, de diferente naturaleza. El archivo *Gothic Past* permite que cualquier usuario añada nuevos monumentos o modifique existentes, por tanto el vocabulario de etiquetas es abierto y dinámico, en contraposición con el vocabulario de *Baroque Art* que era cerrado y por tanto estático. Esto implica además, que el número de personas que intervienen en el etiquetado es muy elevado lo que conlleva gran heterogeneidad en criterios de etiquetado, mientras

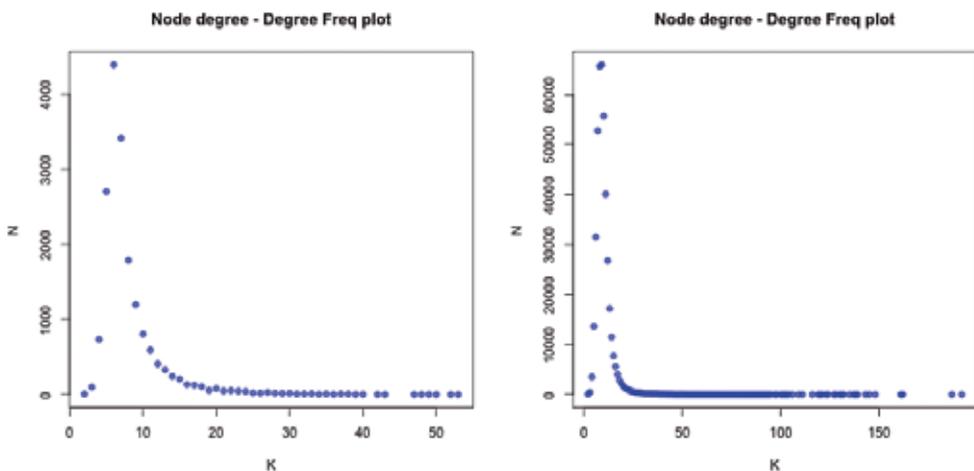


Fig. 4. Distribución de grados de la estructura conceptual asociada al los conjuntos de datos de *Baroque Art* (izquierda) y *Gothic Past* (derecha).

que en el caso del *Baroque Art* el equipo encargado del etiquetado es reducido y organizado. En este caso el conjunto de datos bajo estudio está formado por aproximadamente 3.250 obras y 1.700 etiquetas.

En la figura 4 (derecha) se muestra la distribución de grados de la estructura conceptual asociada al etiquetado de monumentos del gótico en Irlanda, que también cumple la hipótesis SFCH. Vemos que también aparece esa sensible diferencia al inicio de la distribución, por los motivos ya mencionados en el caso del barroco.

Por último es interesante mencionar la diferencia en la complejidad de la estructura conceptual asociada a cada uno de los etiquetados descritos. Aunque en las gráficas de las distribuciones de grados es más difícil observar este hecho, se aprecia claramente en la tabla 2 donde se muestran las principales características de estos conjuntos de datos. Partiendo de contextos formales con dimensiones similares, se deducen de estos retículos de conceptos de tamaño muy diferente. En el caso del etiquetado de obras del barroco, el retículo tiene casi 18.000 conceptos, mientras que en el caso del gótico el retículo tiene casi 420.000 conceptos, esto es, más de veinte veces más.

Esta importante diferencia estructural se debe a la diferente naturaleza de cada etiquetado. Como se ha comentado el etiquetado del barroco usa un vocabulario preestablecido y cerrado en el que intervienen pocos etiquetadores, mientras que en el del gótico el vocabulario es libre y dinámico y en el intervienen todos los etiquetadores que se ofrezcan. Esto hace que el etiquetado del gótico sea mucho más heterogéneo y complejo, si bien, el hecho de que también verifique la hipótesis SFCH indica que es igualmente válido.

Tabla 2. Principales parámetros para cada conjunto de datos. El significado de las columnas se corresponde con las de la tabla 1

	O	A	I	Densidad	CL	<k>
Baroque Art	11.062	221	74.993	3.067%	17.817	7,949
Gothic Past	3.246	1.781	66.432	1,149%	416.896	9,834

CONCLUSIONES Y TRABAJO FUTURO

En este trabajo proponemos el estudio de la estructura de las redes semánticas asociadas a repositorios digitales como medida de la adecuación de su etiquetación. Medir la calidad del etiquetado, una actividad que el usuario o comunidad que construye el repositorio puede hacer con cierto grado de libertad, puede indicarnos cómo de útil es ésta en dos tareas fundamentales: 1) Para facilitar la exploración del repositorio con éxito. 2) Para poder combinar distintos repositorios.

Existe otro aspecto, no tratado en este trabajo por cuestiones de espacio, que consiste en utilizar etiquetados robustos como sistemas de recomendación de etiquetas para los nuevos objetos digitales que deseamos integrar en el repositorio. Si bien los expertos introducen los objetos con etiquetas, el análisis global del repositorio provee de sistemas basados en conocimiento que completarían éste (y, en última instancia, conservaría la estructura del retículo de conceptos). Esta idea, utilizada con éxito en otros campos (Aranda Corral et

al., 2013), sería muy útil cuando el repositorio es suficientemente completo como para extraer conocimiento en forma de sistema experto.

BIBLIOGRAFÍA

- Aranda Corral, Gonzalo A.; Borrego Díaz, Joaquín y Galán Páez, Juan, "Complex Concept Lattices for Simulating Human Prediction in Sport", *Journal of Systems Science and Complexity*, 26 (2013), pp. 117-136.
- Aranda Corral, Gonzalo A.; Borrego Díaz, Joaquín y Galán Páez, Juan, "Scale-free structure in concept lattices associated to complex systems", *Proceedings International Conference on Complex Systems*, 2012a, pp. 1-6.
- Aranda Corral, Gonzalo A.; Borrego Díaz, Joaquín y Giráldez Cru, Jesús, "Conceptual-Based Reasoning in Mobile Web 2.0 by means of Multiagent Systems", *Knowledge Engineering Notes*, 2012b, pp. 176-183.
- Aranda Corral, Gonzalo A.; Borrego Díaz, Joaquín y Giráldez Cru, Jesús, "Agent-mediated shared conceptualizations in tagging services", *Journal of Multimedia Tools and Applications*, 65 (2012c), pp. 5-28.
- Aranda Corral, Gonzalo A. y Borrego Díaz, Joaquín, "Reconciling Knowledge in social tagging web services", *Proceedings International Conference on Hybrid Artificial Intelligence Systems*, 2010, pp. 383-390.
- Au Yeung, Ching-man; Gibbins, Nicholas y Shadbolt, Nigel, "Contextualising tags in collaborative tagging systems", *Proceedings of the 20th ACM conference on Hypertext and Hypermedia*, 2009, pp. 251-260.
- Barabasi, Albert-Laszlo y Albert, Reka, "Statistical mechanics of complex networks", *Reviews of Modern Physics*, 74 (2002), pp. 47-97.
- Carmel, David; Roitman, Haggai y Yom Tov, Elad, "Social bookmark weighting for search and recommendation", *The International Journal on Very Large Data Bases*, 19 (2010), pp. 761-775.
- Clauset, Aaron; Rohilla Shalizi, Cosma y Newman, M.E.J., "Power-law distributions in empirical data", *SIAM Review*, 51 (2009), pp. 661-703.
- Eklund, Peter W. y Wray, Tim, "Social Tagging for Digital Libraries using Formal Concept Analysis", *Proceedings 7th International Conference on Concept Lattices and Their Applications*, 2010, pp. 139-150.
- Ganter, Bernhard y Wille, Rudolf, *Formal Concept Analysis: Mathematical Foundations*, New York, Springer-Verlag, 1997.
- Golder, Scott y Huberman, Bernardo A., "Usage Patterns of Collaborative Tagging Systems", *Journal of Information Science*, 32 (2006), pp. 198-208.
- Halpin, Harry; Robu, Valentin y Shepherd, Hana, "The complex dynamics of collaborative tagging", *Proceedings of the 16th International Conference on World Wide Web*, 2007, pp. 211-220.
- Jäschke, Robert, *Formal Concept Analysis and Tag Recommendation in Collaborative Tagging Systems*, Heidelberg, IOS Press, 2011.
- Jäschke, Robert et al., "Discovering shared conceptualizations in folksonomies", *Journal of Web Semantics*, 6 (2008), pp. 38-53.
- Jung, Jason J., "Knowledge distribution via shared context between blog-based knowledge management systems: A case study of collaborative tagging", *Journal of Expert Systems with Applications*, 36 (2009), pp. 10627-10633.
- Khoo, Michael et al., "Towards digital repository interoperability: the document indexing and semantic tagging interface for libraries (DISTIL)", *Proceedings of the Second international conference on Theory and Practice of Digital Libraries*, 2012, pp. 439-444.
- Lee, Sangjin y Park, Jonghun, "Topic based photo set retrieval using user annotated tags", *Journal of Multimedia Tools and Applications*, 64 (2013), pp. 7-26.

- Motter, E.A. et al. "Topology of the conceptual network of language", *Physical Review E*, 65 (2003).
- Onifade, Olufade; Falade, Williams et al., "Fuzzontology: Resolving Information Mining Ambiguity in Economic Intelligent Process", *Communications in Computer and Information Science*, 54 (2010), pp. 232-243.
- Roget, Peter Mark, *Roget's Thesaurus of English Words and Phrases*, 1911.
- Smith, Gene, *Tagging: People-Powered Metadata for the Social Web*, Safari, Berkeley, New Riders, 2007.
- Steyvers, Mark y Tenenbaum, Joshua B., "The large-scale structure of semantic networks: statistical analyses and a model of semantic growth", *Journal of Cognitive Science*, 29 (2005), pp. 41-78.
- Tanaka, James W. y Taylor, Marjorie, "Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder", *Proceedings of the Fourth International Workshop on Machine Learning*, 1991, pp. 457-482.
- Weick, Karl E.; Sutcliffe, Kathleen M. y Obstfeld, Oscar, "Organizing and the Process of Sensemaking", *Journal of Organization Science*, 16 (2005), pp. 409-421.



RESUMEN

El análisis formal de conceptos (FCA) es una rama de la matemática aplicada cuyo objetivo es el descubrimiento, extracción y organización de conocimiento a partir de datos cualitativos. La teoría provee técnicas para poder razonar con el conocimiento extraído. Es decir, FCA es una aproximación a la minería de datos con una componente semántica (conceptos formales) y una componente lógico-computacional (reglas de asociación y razonamiento automático).

El objetivo de este trabajo es comprobar la aplicabilidad de FCA para tratar conocimiento a partir de repositorios digitales que mantienen la información etiquetada. Desde el punto de vista de FCA, se pretende estudiar la estructura de los conceptos que subyacen en repositorios digitales de arte. La idea es detectar patrones en la estructura conceptual de este tipo de etiquetados, con el objetivo a largo plazo de establecer medidas que estimen la calidad de éstos.

Entre otros, analizaremos etiquetados como los de *Baroque Art* del laboratorio *CulturePlex* (<<http://baroqueart.cultureplex.ca/>>) y el *Visual Archive of Gothic Architecture and Sculpture in Ireland* (<<http://www.gothicpast.com/>>). Compararemos su estructura con la conceptualización asociada a subconjuntos notables de WordNet (<<http://wordnet.princeton.edu/>>).

Bajo la denominada *hipótesis de la conceptualización libre de escala* (Aranda Corral et al., 2012a), se estima la potencia de los etiquetados. También se presentarán técnicas automatizadas para enriquecer el etiquetado mediante sugerencias basadas en razonamiento automático adaptación de las utilizadas en *Delicious* (Aranda Corral et al., 2012b).

Palabras clave: Repositorios digitales, semántica emergente, análisis formal de conceptos, etiquetación de contenidos, humanidades digitales.

ABSTRACT

Formal Concept Analysis (FCA) is a branch of Applied Mathematics whose aim is to discover, extract and organize knowledge from data. FCA provides techniques and tools to reason with such knowledge.

The aim of this paper is to analyse the applicability of FCA to cultural digital repositories that use tagging. It is possible to study the structure of concepts implicit in the

tags by means of FCA, in such way that it is possible to detect some patterns allowing the estimation of the soundness of these tag set.

Among other examples, we apply results from (Aranda Corral *et al.*, 2012a) to analyse two repositories about art: *Baroque Art* from *CulturePlex* Lab (<http://baroqueart.cultureplex.ca/>) and the *Visual Archive of Gothic Architecture and Sculpture in Ireland* (<http://www.gothicpast.com/>). The conceptual structure extracted from both repositories is compared with the one from WordNet (<http://wordnet.princeton.edu/>), which is an example of successful semantic representation.

Under the called *Scale-Free Conceptualization hypothesis* (SFCH) (Aranda Corral *et al.*, 2012a), the soundness of tagging sets (folksonomies) is estimated.

Keywords: Digital Repositories, Emergent Semantics, Formal Concept Analysis, Content tagging, Digital Humanities.

Red ARACNE: retos y objetivos de un proyecto de coordinación en letras hispánicas digitales

CONSOLACIÓN BARANDA
Instituto Universitario Menéndez Pidal (Universidad Complutense)
cbaranda@ucm.es

EDUARDO RODRÍGUEZ
Universidade da Coruña
eduardo.rodriguez.lopez@udc.es

Los especialistas en humanidades hemos acogido con gran entusiasmo las posibilidades que ofrecen las herramientas digitales para la investigación en nuestras áreas de trabajo. Como consecuencia de ello se ha producido tal multiplicación de páginas web relacionadas con la filología que nos puede conducir a un efecto perverso: la pérdida de cada uno de los proyectos en el océano de la red, la inanidad de cada esfuerzo particular.

De la conciencia de este problema surgió la organización de un seminario en junio del año pasado: el *I Seminario Internacional sobre Bibliotecas digitales y Bases de datos especializadas para la investigación en Literaturas Hispánicas (BIDESLITE)* que se realizó en Madrid los días 4 y 5 de julio de 2011. Lo dirigió Mercedes Fernández Valladares y lo organizó el *Grupo de Estudios de Prosa hispánica bajomedieval y renacentista* de la Universidad Complutense de Madrid con financiación del MICINN por medio de una Acción Complementaria (FFI2011-12620-E) y la ayuda del *Instituto Universitario Menéndez Pidal*. Las actas del *Seminario* están publicadas y disponibles en acceso abierto a través de *E-Prints*, el repositorio institucional de la Universidad Complutense <<http://eprints.ucm.es/21207/>>.

Las conclusiones de este seminario se podrían resumir en tres puntos:

- La necesidad de potenciar las estrategias de visibilidad y difusión de los recursos generados al amparo de este tipo de trabajos.
- Convertir esa visibilidad en elemento de identificación y desarrollo de líneas de investigación con base tecnológica de calidad, fortaleciendo la coordinación y su incidencia en el sistema público de investigación.
- La conveniencia de generar instrumentos de enlace y mejora de la recuperación de datos a partir de las bases y bibliotecas digitales existentes. El desarrollo de recursos comunes y herramientas de diseño específico con objeto de proporcionar información de calidad para crear otras herramientas en el área.

La confluencia de estas ideas dio lugar a la percepción, compartida unánimemente entre los ocho grupos participantes, de estar trabajando en proyectos que, aun siendo diferentes en contenidos, objetivos, metodología, géneros literarios o períodos temporales, habían tenido que solventar problemas muy parecidos ofreciendo soluciones –muy dispares en algunos casos– de las que convendría extraer un rédito común.

Como resultado del intercambio de reflexiones se presentó una solicitud de proyecto del conjunto de grupos coordinados en forma de Ayuda Complementaria a 30 meses, con el Dr. Pedro Ruiz como I.P. Estas son las bases de la Red Aracne que presentamos ahora, en la que se integran 6 de los ocho grupos que participaron en el Seminario. Por razones técnicas, los equipos de *CANCIONERO VIRTUAL* de la Universidad de Liverpool y *NICANTO* de la Universidad de Burdeos III no pudieron incorporarse a esta solicitud¹.

La red Aracne está integrada por los siguientes proyectos de investigación:

BIESES. BIBLIOTECA DE ESCRITORAS ESPAÑOLAS (<http://www.bieses.net/>)

En la base de datos de BIESES se recogen todas las referencias primarias y secundarias de las escritoras españolas (en sentido amplio) hasta 1800; el arco temporal enlaza así con la bio-bibliografía de escritoras del siglo XIX de Carmen Simón Palmer.

La información en BIESES está construida y seleccionada en función de una metodología de estudio específica y, aunque útil en otros aspectos y para otros intereses, en un futuro deberá ser capaz de ofrecer los datos necesarios para responder a las preguntas que se formulan desde esa opción crítica. Algunas de estas preguntas implícitas pueden ser las siguientes. En cuanto a la identidad biográfica o social de las autoras: cuántas escritoras hay en el período, cómo se distribuyen cronológica y geográficamente, qué coincidencias vitales hay entre ellas que puedan sugerir posibles conexiones o relaciones; cuántas de estas escritoras pertenecieron a órdenes religiosas, qué órdenes, en qué conventos, cuándo. En lo que se refiere a sus obras: cuántas y de qué tipo están en manuscritos, qué se edita, dónde, cuándo, cómo, por medio de qué agentes culturales o mediadores. Por último, BIESES aún puede responder a otras cuestiones de recepción, a través del campo de las menciones coetáneas que recogen citas de otros autores, de las cartas intercambiadas, de los paratextos o del intercambio poético; también se revisan inventarios de bibliotecas con el fin de averiguar qué autoras son leídas y por quién o por qué autoras que se espera sean citadas no lo son en algunos contextos.

BIDISO. BIBLIOTECA DIGITAL SIGLO DE ORO (<http://www.bidiso.es/sielae/>)

El grupo SIELAE (Seminario Interdisciplinar para el Estudio de la Literatura Áurea Española) ha propiciado desde 1992 la creación sucesiva de varias bases de datos y bibliotecas digitales en torno a cuatro proyectos:

¹ Aunque por razones históricas *NICANTO* no ha podido realizar todavía el desarrollo informático, está proporcionando material inédito y desconocido o poco conocido a algunos grupos de la red-Aracne; es ejemplo de trabajo útil y fundamental, a pesar de no contar con herramientas informáticas actuales.

- Emblemática.
- Relaciones de sucesos.
- Polianteas y libros de erudición.
- Inventarios y Bibliotecas del Siglo de Oro (IBSO).

El proyecto se presenta en este *I Congreso Internacional de la HDH*, por lo que remitimos a su intervención.

BSF. BIBLIOTECA SAAVEDRA FAJARDO DE PENSAMIENTO POLÍTICO HISPÁNICO (<http://saavedrafajardo.um.es>)

La labor fundamental de la Biblioteca Saavedra Fajardo es centralizar una serie de herramientas útiles para el investigador dedicado al pensamiento político hispano, así como presentar un catálogo extenso y materiales adicionales de forma ordenada, para acceder con facilidad a recursos propios y externos. Es, en este momento, la que cuenta con mayor volumen de información especializada de nuestra asociación.

Atiende al estudio de las fuentes hispánicas y a la producción filosófica en nuestro idioma y está integrada por cuatro colecciones, una base de datos bio-bibliográfica de fuentes primarias y secundarias sobre pensadores hispánicos, una hemeroteca digital con 11 títulos de revistas especializadas y una sección de Tribuna con artículos breves y pequeños ensayos, abierta a la colaboración externa. Ofrece documentación biográfica y bibliográfica, edita fuentes de la historia del pensamiento hispánico, en versión de fichas, con índices descriptivos de cada página, y en versión pdf, con textos completos, y pone en la red artículos y trabajos de investigación sobre filosofía y pensamiento español de todas las épocas.

El Centro de Documentación Gonzalo Díaz y María Dolores Abad vertebró el funcionamiento de la BSF; es un ingente repertorio biobibliográfico de pensadores de la filosofía española que tiene como núcleo la magna obra de Gonzalo Díaz *Hombres y documentos de la filosofía española*.

A día de hoy hay publicados en el GDDA los repertorios de 2.666 pensadores hispanos, con más de 52.000 obras y más de 33.000 estudios referenciados. Cuando una obra o estudio está disponible en Internet, basta con seleccionar la referencia en el Centro de Documentación para acceder a la misma.

CLARISEL. BASES DE DATOS (<http://clarisel.unizar.es/>)

Es un portal que reúne tres bases independientes.

Los trabajos incluidos en *AMADÍS* pretenden arrojar nuevas luces sobre la producción literaria caballeresca compuesta en diferentes modalidades genéricas desde la Edad Media hasta el siglo XVII, aparte de incluir estudios posteriores que tengan como referencia intertextual obras anteriores.

Además, si bien el ámbito de trabajo es fundamentalmente filológico, se incorporan entradas interdisciplinarias –de historia, derecho, arte, folclore, antropología, etc.–, que puedan ayudar a interpretar y contextualizar la materia objeto de atención en un ámbito ibérico. Esto implica que incluye referencias temáticas, como los trabajos sobre los “rieptos” o sobre tácticas militares, órde-

nes militares, etc., referidas a textos castellanos, catalanes, gallegoportugueses y portugueses. El término de partida es 1998 por ser la fecha de los inicios oficiales de la página, teniendo en cuenta, además, que en el 2000 se publicó la bibliografía sobre los libros de caballerías de Daniel Eisenberg y M^a Carmen Marín.

SENDEBAR recoge y sintetiza la bibliografía sobre el cuento medieval hispánico a partir de 2005. Incluye ediciones y estudios sobre cuentos escritos en castellano, catalán y gallego-portugués, así como en otras lenguas occidentales y orientales (latín, árabe, hebreo, portugués, francés, italiano, etc.), siempre que se trate de obras que mantengan conexiones, directas o indirectas, con producciones hispanas. Con idénticas restricciones, se incorporan publicaciones sobre el relato folclórico y sus métodos de análisis, dada la interrelación entre el cuento escrito en la Edad Media y la tradición oral.

HEREDIA está dedicada a la literatura escrita en Aragón durante la Edad Media y el siglo XVI, con independencia también de la lengua; así, en ella se incorpora información sobre Pedro Alfonso, literatura aljamiado-morisca, escritos en aragonés o en catalán, etc. Como la anterior, fue creada en el año 2005, si bien surgió ya con la vocación de incluir, poco a poco, referencias impresas anteriores; cuenta en la actualidad con más de 1.230 entradas.

En todos los casos, además de bibliografía y resúmenes de contenido, se ofrecen los enlaces electrónicos que están disponibles tanto a los textos como a las revistas mencionadas.

DIALOGYCA BDDH. BIBLIOTECA DIGITAL DE DIÁLOGO HISPÁNICO (<http://www.ucm.es/info/dialogycabddh/>)

Se propone recopilar, analizar y poner a disposición de los investigadores el *corpus* de diálogos hispánicos, desde sus orígenes hasta la actualidad, mediante una base de datos y una biblioteca digital. De acuerdo con estos planteamientos, hemos elaborado una base de datos que proporciona información sobre cada uno de los diálogos, ahora preferentemente de los siglos XV a XVII, pues en esta primera fase nos hemos concentrado en la etapa dorada del género, aunque eso no es obstáculo para introducir fichas analíticas de otros períodos.

Atendemos a los siguientes aspectos: el autor y otros responsables secundarios de la obra, las características literarias y modalidades de difusión, los personajes, las materias principales, los temas y formas de enunciación y lengua, la bibliografía primaria –con la localización y referencia de los testimonios manuscritos e impresos– y la bibliografía secundaria específica sobre cada diálogo.

Pero además, *Dialogyca BDDH* ofrece la reproducción completa del testimonio más relevante para su transmisión –de ahí el carácter de Biblioteca o colección digital–, en unos casos –hasta ahora la mayoría– digitalizado por nosotros y presentado en un documento pdf con permiso de difusión de la biblioteca que lo custodia; en otros mediante enlaces a las digitalizaciones realizadas por otras instituciones, si las hay.

De este modo, cada ficha encierra de forma organizada un estado de la cuestión sobre un diálogo concreto junto con su texto, y la Biblioteca digital proporcionará una visión de conjunto sobre el género: permitirá conocer mejor el número de autores y obras, ayudará a percibir la desigual atención crítica recibida por cada uno de ellos y favorecerá la posibilidad de establecer redes

de relaciones sincrónicas y diacrónicas entre los textos, sus temas, sus lugares o fechas de impresión, sus personajes, sus fuentes, etcétera.

Este diseño, basado en el concepto de *obra*, marca una diferencia sustancial con otras bases de datos de tipo catalográfico o bibliográfico y de carácter generalista, cuya unidad articuladora suele ser el documento –manuscrito o impreso– y atienden preferentemente al lugar o institución en que se ha conservado; por el contrario, el concepto de *obra* como unidad articuladora nos permite agrupar testimonios y documentos de diversa tipología, conservados en bibliotecas muy dispersas y recoger de ese modo, junto a la información literaria la tradición textual de cada diálogo.

PHEBO. POESÍA HISPÁNICA EN EL BAJO BARROCO (<http://phebo.es>)

El proyecto *PHEBO* (Poesía Hispánica en el Bajo Barroco) surgió a mediados del 2010 en el seno del equipo que desarrollaba el estudio “La poesía de bajo barroco: repertorio y categorías”. El objetivo que se plantea es sentar las bases para una cartografía de la poesía de este período, tan poco atendido por la crítica y la historia literarias, a partir de dos elementos básicos: un repertorio de los impresos con materia poética entre 1650 y 1750, de un lado, y una revisión de los conceptos y categorías aplicados al análisis de la poesía áurea, de otro. La base de datos persigue el establecimiento de las herramientas básicas para el análisis, estudio y valoración de la práctica poética en el Bajo Barroco mediante la elaboración del repertorio de impresos poéticos entre 1650 y 1750, la edición de los textos y la recopilación de la bibliografía secundaria especializada. El elemento distintivo lo constituye la edición, esto es, la práctica fundamental de la filología y la base de todo estudio que pueda ser compartido.

Se decía al comienzo que somos conscientes de las diferencias en contenidos, objetivos, metodología, géneros literarios o períodos temporales de nuestros proyectos, pero también lo somos de que, en parte por ello, nos complementamos y también de que compartimos puntos de partida importantes. Como objeto de estudio, predominan los proyectos que ambicionan ofrecer un panorama de conjunto sobre un género literario: Cuento medieval, Libros de caballerías, Diálogo Literario, Emblemática, Relaciones de sucesos, Polianteas o libros de erudición y Poesía del bajo barroco.

Otros integran más de un género, pues han partido de diferentes criterios y tienen un carácter transversal: *BIESES*, *Heredia*, *IBSO* y *Saavedra Fajardo*. El abanico temporal se extiende desde la edad media hasta hoy, con especial presencia de los siglos XV, XVI y XVII en este momento.

Cada uno de los proyectos tiene objetivos claramente delimitados, ofrece información discriminada y organizada de acuerdo con esos objetivos, aporta valor añadido a sus respectivos ámbitos de estudio, recoge información altamente especializada y busca la utilidad pública, así como servir de referencia informativa en su campo. Todos integran el contexto de un modo u otro, recogen fuentes primarias y, en muchos casos, también secundarias; aunque trabajan fundamentalmente sobre impresos, también se encuentran presentes los manuscritos. Comparten también un sentido abarcador de las Humanidades Hispánicas, pues incluyen textos en las distintas lenguas peninsulares y también hispano-latinos, aljamiados, portugueses, etcétera.

Se trata de proyectos en marcha, con procesos de actualización permanentes, que han previsto la colaboración de investigadores en formación, junto con especialistas y colaboradores externos, de ahí que consideren importante hacer constar la autoría de cada una de las aportaciones, elemento fundamental como respaldo del rigor científico. Por supuesto, cuentan con financiación pública y con un soporte institucional que auspicia su estabilidad.

La Red Aracne se apoya en dos pilares: el carácter complementario de los proyectos que la integran y la concurrencia de criterios en el planteamiento científico de todos ellos. El objetivo último no es sumar la información, sino multiplicarla, porque se trata de poner en relación datos previamente procesados, elaborados y estructurados con rigor². Pretendemos, por una parte, rentabilizar los medios y el trabajo invertidos potenciando su utilización y conseguir visibilidad, pero también construir una vía de futuro inédita para las Humanidades Digitales Hispánicas. La red Aracne podrá servir como banco de pruebas y futuro punto de convergencia de otros grupos que trabajan en el campo de las Humanidades Digitales.

Para la creación de la red Aracne hemos elegido avanzar con dos núcleos de trabajo fundamentales:

1. Creación de un portal web que centralice y ordene la información sobre estos proyectos e instrumentos en el campo de las humanidades digitales aplicadas a las letras hispánicas, con objeto de que sirva de referencia, orientación, convergencia y principio de coordinación entre iniciativas, las ya existentes y las emergentes.
2. Paralelamente, la elaboración de un *metabusca* que ofrezca la posibilidad de una búsqueda simultánea en todas las plataformas participantes y ofrezca resultados coherentes y exhaustivos, aunque no necesariamente completos, de todas las informaciones que aparezcan en las distintas bases de datos. Este tipo de experiencia y las problemáticas, tanto técnicas como metodológicas, que se lleguen a identificar y resolver, servirán para crear pautas para la incorporación de nuevos proyectos a la red y también como ejemplo para los proyectos que quieran compartir los resultados, pero no la aplicación informática.

Además, de manera paralela o sucesiva, según los casos, nos proponemos avanzar en otros objetivos:

- Diseño de un modelo de criterios de evaluación de calidad técnica y científica de los productos resultantes, con la finalidad de homogeneizar protocolos y propiciar su aceptación para el reconocimiento y evaluación de los trabajos realizados.
- Desarrollo de una herramienta flexible de carácter interactivo orientada a la enseñanza y formación de los alumnos en los inicios de la carrera investigadora.

² A modo de prueba, Mercedes Fernández Valladares hizo una simulación buscando en *Dialogyca BDDH* datos relativos a Aragón que pudiesen complementar la base *Heredia* y la variedad de resultados fue superior a la que se podía prever: aparecieron autores que nacieron, murieron o desempeñaron actividades profesionales en Aragón, libros allí impresos o conservados en sus bibliotecas, e incluso personajes, interlocutores de algún diálogo, estaban relacionados con Aragón.

- Desarrollo de la ampliación del proyecto y búsqueda de convocatorias oficiales y fuentes de financiación (en especial en el marco europeo), para llevarlo a cabo.

Con ellos se busca establecer una sinergia que permita la confluencia de los equipos que inicialmente respaldan el proyecto más aquellos que en el futuro deseen sumarse a partir de la colaboración en actuaciones concretas y delimitadas que permitan una homologación de estrategias de investigación, una integración de resultados y una tecnología común.

En lo referente a los desarrollos tecnológicos, cuya responsabilidad ha sido asignada al Laboratorio de Bases de Datos de la Universidad de A Coruña (<http://lbd.udc.es>), en la actualidad se están avanzando en las tareas conducentes al cumplimiento de los dos objetivos fundamentales de proyecto, que son, la elaboración de un portal web para la difusión de la *Red* y su actividad, y la implementación de una herramienta de búsqueda que permita buscar en todas las bibliotecas digitales y bases de datos integradas en la misma.

Metabuscador de recursos digitales

El principal problema con el que nos encontramos en el desarrollo del *metabuscador* es la heterogeneidad de los repositorios digitales a integrar. Cada biblioteca digital o base de datos que forman parte la *Red* ha sido diseñada atendiendo a unos criterios específicos, propios del dominio de los recursos con los que trabaja, e implementada utilizando diferentes herramientas software que, además, muchas veces condicionan notablemente la propia estructuración de los datos catalogados. Además, la integración ha de hacerse de forma que no se modifiquen las bibliotecas digitales y bases de datos originales.

Este problema de la heterogeneidad no es específico del proyecto *Red ARACNE*, sino que afecta en mayor o menor medida a todos los proyectos de integración de colecciones documentales digitales. Existen diversas aproximaciones para construir sistemas que permitan integrar distintas fuentes de datos de cara a su interoperabilidad. La solución que mejor se adapta a las necesidades concretas de la *Red ARACNE* es la de utilizar el protocolo OAI-PMH (*Open Archives Initiative-Protocol Metadata Harvesting*, <http://www.openarchives.org/pmh/>) para la compartición de los metadatos descriptivos de los recursos digitales, sobre los que posteriormente se resolverán las búsquedas. Al tratarse de un estándar abierto y ampliamente utilizado en la actualidad por distintas iniciativas institucionales, la utilización del protocolo OAI-PMH permitirá también la difusión de los recursos digitales de la *Red ARACNE* a través de grandes repositorios de recursos digitales como Europeana (<http://www.europeana.eu>), promovido por la Unión Europea, o Hispana (<http://hispana.mcu.es>), del Ministerio de Educación, Cultura y Deporte.

Los recursos digitales se describirán utilizando el conjunto de 15 elementos básicos Dublin Core (*Dublin Core Metadata Initiative*, <http://dublincore.org/>), entre los que se encuentran el título, la autoría, una descripción, la temática, el idioma o el formato. Todos estos elementos son opcionales, lo que permite una gran flexibilidad a la hora de describir a los recursos aun cuando la información que se dispone es parcial o incompleta.

Por tanto, el primer paso ha sido el efectuar un estudio pormenorizado de los repositorios digitales que forman parte de la *Red*, analizando la información

que almacenan y cómo se puede establecer la correspondencia entre estos datos y los elementos básicos de Dublin Core. Este análisis se ha visto dificultado por la ausencia de documentación técnica (como documentos de requisitos o de diseño) de los repositorios, y también por el hecho de que algunos de ellos han sido desarrollados utilizando aplicaciones informáticas propietarias, no abiertas, lo que dificulta el acceso a la información que almacenan.

Como resultado de este análisis se ha concluido que todas las bibliotecas digitales y bases de datos pertenecientes a los grupos que forman parte de la *Red ARACNE* van a poder ser integradas en el *metabusador*. En cualquier caso, para cada una de ellas será preciso diseñar e implementar un sistema mediador (o *proxy*) específico, que permita traducir el formato de datos propio de la biblioteca digital al vocabulario común Dublin Core. La información obtenida en esta fase de análisis resulta de gran utilidad también para depurar los criterios tecnológicos establecidos para la adhesión de una nueva biblioteca digital o base de datos a la *Red*, y que se publicarán en el portal web *Red ARACNE*.

Portal Red ARACNE

En la actualidad se dispone ya de una primera versión pública del portal Red ARACNE, accesible en <http://www.red-aracne.es>. A continuación se describen brevemente las distintas secciones de las que se compone:

1. **Presentación.** Contiene información básica de la *Red*, como la relativa a su motivación original o sus objetivos.
2. **¿Qué ofrecemos?** Ofrece información general sobre lo que puede encontrarse en la *Red ARACNE*, y particular sobre los grupos de investigación, las bibliotecas digitales y las bases de datos especializadas que forman parte de ella.
3. **Publicaciones.** Ofrece una recopilación de las publicaciones relevantes en el ámbito de la *Red*, organizadas en tres secciones: revistas, bibliografía y biblioteca.
4. **Noticias.** Muestra una recopilación de noticias sobre la *Red ARACNE* o cualquier otra información relevante dentro del ámbito de las Humanidades Digitales.
5. **Enlaces.** Ofrece una lista de enlaces a otros espacios web relacionados con el proyecto.
6. **Contacto.** Incluye un formulario que permite contactar con los responsables de la *Red*.

Una sección básica del portal *Red ARACNE*, y que todavía no está disponible, será la interfaz pública del *metabusador*, que permitirá buscar recursos de forma simultánea en todas las bibliotecas digitales y bases de datos integradas en la *Red*.

El portal web *Red ARACNE* incluye una herramienta de administración privada (protegida por contraseña) que permite la gestión de toda la información disponible en él. A esta herramienta de administración pueden acceder tanto los responsables de la gestión de los contenidos de la *Red*, como los responsables de la gestión de los contenidos de cada grupo, si bien estos últimos tendrán acceso restringido únicamente a la información propia de su grupo y de sus bibliotecas o bases de datos.

Por último, destacar que en el desarrollo del portal se ha seguido una arquitectura software modular, a fin de facilitar las más futuras ampliaciones tanto de nuevas secciones como de nuevos grupos y/o bibliotecas digitales³.



RESUMEN

La red-Aracne ofrece una novedad importante: la coordinación de seis proyectos de investigación en Humanidades Digitales (BIESES, BIDISO, BIBLIOTECA SAAVEDRA FAJARDO, CLARISEL, DIALOGYCA BDDH y PHEBO). Se comentan los distintos puntos de partida y los retos que deben abordar para conseguir una coordinación que persigue rentabilizar el trabajo de cada uno de ellos, conseguir más visibilidad en la red y facilitar el acceso de los usuarios a un volumen de información cuantioso y organizado.

Palabras clave: Red de proyectos de investigación, metabuscador.

ABSTRACT

The red-Arachne offers an important novelty: the association of six research projects in Digital Humanities (BIESES, BIDISO, LIBRARY SAAVEDRA FAJARDO, CLARISEL, DIALOGYCA BDDH and PHEBO). We discuss the different starting points and the challenges that must be addressed to achieve coordination in order to optimize the work for each of them, for greater visibility in the web and facilitate users access to a large volume of organized information.

Keywords: Research projects network, Meta searcher.

³ Trabajo realizado en el marco del Proyecto: Inventario, Descripción, Edición Crítica y Análisis de Textos de Prosa Hispánica Bajomedieval y Renacentista. Línea Diálogos (Fase 3) FFI2012-33903 y de la Acción complementaria FFI2011-15606-E (subprograma FILO).

Textos antiguos vascos en Internet¹

GIDOR BILBAO

Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)

gidor.bilbao@ehu.es

RICARDO GÓMEZ

Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)

ricardo.gomez@ehu.es

1. TEXTOS ANTIGUOS VASCOS

Conviene comenzar este trabajo definiendo lo que entendemos como *texto antiguo vasco* y recordando algunas características del corpus al que nos referimos.

La histórica y aún vigente recopilación de textos vascos de Luis Michelena, publicada en 1964, recogía textos breves escritos en lengua vasca antes del año 1700, y se llamó *Textos Arcaicos Vascos*. Actualmente, en los estudios de historia de la lengua vasca se acepta comúnmente que el año 1745, fecha de publicación del *Diccionario Trilingüe del castellano, bascuence y latín* de Manuel Larramendi, marca el final del euskera antiguo y el comienzo del llamado "primer euskera moderno" (Lakarra, 1997: 516). La periodización propuesta por Lakarra cuenta hoy en día con una amplia aceptación y es la utilizada en la monografía más importante sobre la historia del euskera (Gorrochategui, Igartua y Lakarra, 2013). Por ello hemos preferido utilizar el término *texto antiguo vasco*, refiriéndonos así a todo texto, no necesariamente literario, de diversa extensión, escrito en lengua vasca antes de 1745.

Para el investigador menos familiarizado con los textos vascos, quizás no resulta superfluo recordar aquí que, a pesar de la amplitud del período abordado, el corpus es muy reducido. Sirva como referencia un estudio de Ibon Sarasola sobre la producción editorial vasca en el período antiguo (1976: 177-183), que, a pesar del tiempo transcurrido desde su publicación, aún sigue siendo significativo; recogemos en una tabla de elaboración propia algunos datos que nos parecen relevantes, referidos en este caso a publicaciones de más de 48 páginas:

¹ Los autores han contado con la ayuda de estos proyectos y grupos de investigación: "Monumenta Linguae Vasconum (IV): textos arcaicos vascos y euskera antiguo" (MINECO, FFI2012-37696), "Historia de la lengua vasca y lingüística histórico-comparada" (HLMV-LHC) (Gobierno Vasco, GIC. IT698-13) y "Lingüística Teórica y Diacrónica: Gramática Universal, lenguas indoeuropeas y lengua vasca" (LingTeDi) (UPV/EHU, UFI11/14).

Período	Publicación total	Producción total (sin reediciones)	Producción original (sin traducciones)	Producción original NO religiosa
1545-1599	3	3	1	1
1600-1649	11	7	6	-
1650-1699	25	13	6	1
1700-1749	34	17	5	1

Actualmente no disponemos de datos exactos, con número de caracteres, del corpus que nos ocupa atendiendo a la definición que hemos adelantado más arriba; pero los datos referidos a la producción editorial son perfectamente extrapolables, de modo que podemos afirmar que los textos en lengua vasca de la Antigüedad y de la Edad Media son escasísimos y se reducen a nombres propios de persona, de divinidades o de lugar, a los que se unen, a partir de la Edad Media, glosas, vocabularios y frases, expresiones o palabras insertas en documentos latinos y romances. En los siglos XVI y XVII, encontramos esta misma tipología, aunque en mayor cantidad, pero aparecen también algunas declaraciones en procesos judiciales, cantares o fragmentos de cantares antiguos, poesías sueltas, colecciones de refranes y sentencias, métodos para el aprendizaje de la lengua vasca, y las primeras obras de cierta extensión (el primer libro impreso en euskera, *Linguae Vasconum Primitiae* de Bernard Etxepare, data de 1545). La progresión es notable a partir de principios del siglo XVIII, pero siempre en cantidades tan ínfimas que cualquier texto escrito en lengua vasca antes de 1745, por pequeño y aparentemente insignificante que pueda parecer, es importante para la historia de la lengua vasca.

2. EL ACCESO A REPRODUCCIONES DE ORIGINALES: BIBLIOTECAS DIGITALES

Los facsímiles de textos antiguos vascos, tanto manuscritos como impresos, accesibles desde internet son, como sucede con otras lenguas del entorno, cada vez más numerosos.

Las bibliotecas digitales y los proyectos internacionales de tipo generalista son, por supuesto, útiles para este fin. Por ejemplo, gracias al proyecto *Europeana*², podemos acceder en menos de un minuto, a través de *Gallica*³, a una reproducción de calidad excelente del *unicum* de la primera edición de la obra de Bernard Etxepare (1545) que hemos mencionado más arriba, para cuya consulta debíamos acudir hasta hace bien poco a París, a la Biblioteca Nacional de Francia. Desde esa misma plataforma *Europeana*, necesitaremos un minuto más para poder consultar otra reproducción ejemplar del manuscrito *Grammaire cantabrique* de Pierre d'Urte (c. 1714), disponible en los fondos de la Fundación Sancho el Sabio de Vitoria-Gasteiz, a través de su proyecto *Memoria Digital Vasca*⁴, integrado en *Hispana*⁵. Otros proyectos como Google Books o Internet Archive nos ayudan a diario en nuestras búsquedas.

² <<http://www.europeana.eu/>>.

³ <<http://www.gallica.bnf.fr/>>.

⁴ <<http://www.memoriadigitalvasca.es/>>.

⁵ <<http://hispana.mcu.es/>>.

También las bibliotecas digitales locales y especializadas en textos vascos han apostado en los últimos años por permitir el acceso digital a facsímiles de ediciones impresas y manuscritos de textos antiguos vascos. Por razones de espacio, sólo podemos mencionar las más importantes⁶:

- **Liburuklik – Biblioteca Digital Vasca**⁷: reúne y facilita el acceso a parte de los recursos desarrollados en la Comunidad Autónoma del País Vasco, entendiendo el patrimonio bibliográfico vasco en su sentido más amplio (patrimonio bibliográfico existente en bibliotecas de la Comunidad Autónoma Vasca, independientemente de su origen o su lengua). Ha integrado fondos tan importantes como los Fondos Digitales del Parlamento Vasco (en su totalidad), la biblioteca del Santuario de Loyola, Eresbil (archivo de la música vasca), o la Fundación Sancho el Sabio, entre otros, pero aún no ha conseguido convertirse en referente único, puesto que hay instituciones importantes que no se han integrado en el proyecto (es el caso del Repositorio Digital de la Diputación Foral de Gipuzkoa), e incluso las instituciones que participan lo hacen, en algunos casos, con parte de sus fondos, y no con la totalidad de su patrimonio catalogado, de modo que el investigador se ve obligado a repetir la búsqueda en el catálogo de la propia institución o incluso en otros catálogos más generales.
- **Memoria Digital Vasca**⁸: La Fundación Sancho el Sabio, especializada en estudios vascos, fue una de las instituciones pioneras en la publicación digital de reproducciones de textos antiguos vascos. Se ha integrado en *Liburuklik* y en *Hispana*, pero mantiene su catálogo exclusivo, que sigue siendo la única fuente de información totalmente segura para acceder a la totalidad de sus fondos digitales.
- **Repositorio Digital de la Diputación Foral de Gipuzkoa**⁹: Es una institución pública local que conserva, entre otros, el Fondo Julio Urquijo, de modo que incluye un rico patrimonio de textos vascos, impresos y manuscritos.
- **Biblioteca Navarra Digital**¹⁰: Tiene como finalidad difundir el patrimonio bibliográfico de Navarra a través de las nuevas tecnologías de la información, de tal forma que incluye las reproducciones de varios ejemplares de textos antiguos que difícilmente podrían ser consultados de otra manera.
- **Biblioteca Foral de Bizkaia. Biblioteca Digital**¹¹: La Diputación Foral de Bizkaia es heredera de diversos legados públicos y privados que incluyen textos antiguos vascos impresos y manuscritos, cuyos facsímiles están siendo publicados en internet paulatinamente en las colecciones “Manuscritos vascos” y “Fondo antiguo vasco”, fundamentalmente.
- **Biblioteca Azkue de Euskaltzaindia**¹²: La biblioteca de la Academia de la Lengua Vasca es depositaria, entre otros, de los fondos de la Fundación

⁶ Dejamos al margen los sistemas de consulta de la documentación de los archivos (valga como ejemplo el sistema *Badator* de la Comunidad Autónoma de Euskadi [<http://dokuklik.snae.org/>], cuya labor de los últimos años hace accesibles desde internet documentos que contienen textos antiguos vascos inéditos [Bilbao, 2013]).

⁷ <http://www.liburuklik.euskadi.net/>.

⁸ <http://www.memoriadigitalvasca.es/>.

⁹ <http://meta.gipuzkoakultura.net/>.

¹⁰ <https://administracionelectronica.navarra.es/binadi/busqueda.aspx>.

¹¹ <http://bibliotecaforal.bizkaia.net/>.

¹² <http://gorbeia.euskaltzaindia.net/cgi-bin/abwebp/LLeu-ES/G0/>.

Olaso Dorrea, una de las colecciones más ricas de textos vascos antiguos. Se puede consultar el catálogo en internet, y las reproducciones pueden ser consultadas en CD-ROM, en la sala de la biblioteca.

Finalmente, hay reproducciones de originales de textos vascos disponibles en bibliotecas digitales aisladas, en las cuales el estudioso difícilmente buscaría si no es porque tiene conocimiento fehaciente de su existencia. Es el caso, por ejemplo, de los manuscritos del siglo XVII, conocidos como *Basque-Icelandic Glossaries*, que han circulado entre los filólogos interesados gracias a unas copias microfilmadas a mediados del siglo XX, de calidad aceptable pero sin punto de comparación con las fotografías que actualmente podemos consultar en nuestro ordenador, en la página web de *The Árni Magnússon Institute for Icelandic Studies*¹³.

Desgraciadamente, el investigador interesado en estudiar textos vascos antiguos debe necesariamente, en cada caso, conocer o indagar en las diferentes posibilidades, puesto que, a pesar de tratarse de un corpus reducido, no existe aún un directorio especializado y exhaustivo que permita localizar fácilmente los cada vez más numerosos facsímiles accesibles desde internet, pero dispersos por diferentes bibliotecas y proyectos de ámbito nacional o internacional, generalistas y especializados.

3. EL ACCESO A LOS TEXTOS: FORTALEZAS Y DEBILIDADES DE DIVERSOS PROYECTOS

3.1. *Euskal Testuen Gordailua* [Depósito de textos vascos]

Este portal de textos antiguos vascos¹⁴ se encuentra alojado en la web de la UPV/EHU y su responsable es Patxi Salaberri Muñoa, profesor de la misma institución. El portal tiene como objetivo reunir en un único repositorio versiones electrónicas de textos vascos de diversa procedencia, o bien confeccionadas *ad hoc* a partir de distintas fuentes impresas. Hace años que no se renuevan sus contenidos, siendo la última actualización de 29 de diciembre de 2000.

El portal cuenta con tres secciones:

1. "Idazleak eta idazlanak" [Autores y obras]: son cincuenta y tres obras escritas desde el s. XVI a la primera mitad del XX, organizadas alfabéticamente por autor. Gran parte de los "clásicos" que incluye fueron proporcionados por Josu Lavín, el auténtico precursor en la digitalización de textos vascos.
2. "Eresiak, bertsoak eta kantak" [Elegías, versos y cantos]: se trata de versos antiguos, en su mayor parte tomados de las colecciones publicadas por Luis Michelena (1964) e Ibon Sarasola (1983).
3. "Kutxatila" [Arqueta]: frases sueltas, cartas y otros textos breves, casi todos tomados también de Michelena (1964) y Sarasola (1983).

Junto a cada texto se menciona la versión (o edición) publicada en papel utilizada (bajo el epígrafe "Bertsio honen iturria") o, en su caso, el autor de la

¹³ <http://www.arnastofnun.is/page/basknesk_islensk_ordasofn_en>.

¹⁴ <<http://www.vc.ehu.es/gordailua/>>.

versión electrónica preparada ex profeso para el repositorio y, en ocasiones, la fuente original ("Non"). Se ofrece únicamente la transcripción del texto, sin introducción ni notas de ningún tipo. Por otro lado, no hay un criterio uniforme para la presentación de los textos, conservando las características muy diversas de la fuente impresa o de la versión electrónica utilizada en cada caso.

3.2. Corpus del *Diccionario General Vasco – Orotariko Euskal Hiztegia (OEH)*

Fue un subproducto del *Diccionario General Vasco* de la Academia de la Lengua Vasca, diseñado a partir del corpus elaborado para la confección de dicho diccionario. Fue editado en 2000, en soporte CD-ROM, pero no fue comercializado y tuvo una distribución limitada, como versión de trabajo exclusiva para las diferentes comisiones de la Academia de la Lengua Vasca. No obstante, merece un lugar privilegiado en este trabajo, porque fue pionero en muchos aspectos y buena parte de sus materiales fueron posteriormente utilizados por *Klasikoen Gordailua*.

El CD-ROM incluye 304 documentos, lo cual suponía en el 2000 un volumen más que considerable. Todos los textos incluidos son obras literarias (en su acepción más amplia y generosa, acorde con las limitaciones de la producción en lengua vasca), escritas entre 1545 y 1970, pero quedaron excluidos textos editados por Michelena (1964), y otros textos de pequeño volumen como romances, etcétera.

Este trabajo incorporaba un motor de búsqueda muy rudimentario desde la perspectiva actual, pero novedoso y útil en aquel momento. El investigador puede organizar sus búsquedas con filtros cronológicos, de variedades dialectales o géneros, y puede restringir la búsqueda a una obra concreta. Por supuesto, el buscador no podía incorporar entonces ningún tipo de lematizador, de modo que se buscan palabras enteras (secuencias de caracteres), pudiendo combinar hasta dos palabras, relacionadas con diferentes operadores. En la tabla de resultados obtenidos, que presenta una lista alfabética de palabras, el investigador debe buscar manualmente otras palabras relacionadas con el lema buscado. El motor de búsqueda permite entonces consultar los resultados en su contexto e incluso obtener la obra completa en formato RTF.

Los textos, que como hemos apuntado más arriba, no habían sido preparados para su publicación, sino como material de trabajo para la confección del *Diccionario General Vasco*, son presentados siempre en grafía modernizada y deben ser manejados con cuidado, puesto que la carencia de labor de edición es evidente: la modernización de la grafía no siempre es correcta, no hay notas ni aparato crítico, ni siquiera explicaciones sobre las modificaciones operadas en cada texto. Debemos insistir, sin embargo, en que este trabajo cumplió en su momento una función importante y, usado con sentido crítico, resultó ser una herramienta muy útil para los investigadores.

3.3. *Klasikoen Gordailua* [Depósito de clásicos vascos]

Este segundo repositorio¹⁵ forma parte del más amplio portal *Armiarma* [La tela de araña]¹⁶, dedicado a la literatura vasca, y todos sus contenidos son de

¹⁵ <<http://klasikoak.armiarma.com/>>.

¹⁶ <<http://www.armiarma.com/>>.

dominio público. *Klasikoen Gordailua* es el heredero de *Euskal Testuen Gordailua*, pero supone un gran avance con respecto a su predecesor. Por un lado, se han incorporado nuevos textos, hasta completar el corpus de mayor volumen de textos antiguos vascos (el límite *ante quem* es 1936) accesible hoy en día en internet. Por otra parte, el repositorio dispone de un motor de búsqueda relativamente avanzado, denominado "Corpus arakatzaillea" [El rastreador del corpus]. Permite buscar hasta un máximo de tres palabras (o inicios de palabra) a la vez y establecer la distancia entre ellas. Además, la búsqueda puede restringirse mediante varios filtros: autor/obra, género, época y dialecto. Un problema que plantea este motor de búsqueda es que sólo funciona dentro de la sección de autores y obras, a la que nos referiremos a continuación, pero no en el resto de los textos recogidos en el repositorio.

El repositorio cuenta, en efecto, con tres secciones, que se corresponden, corregidas y muy aumentadas, a las tres de *Euskal Testuen Gordailua* que hemos mencionado en la sección anterior.

En primer lugar, "Idazleak eta idazlanak" [Autores y obras], que constituye el corpus principal del repositorio. Esta sección consta en la actualidad de 496 obras, 11,9 millones de palabras y 73,9 millones de caracteres. Puede descargarse completa en formato TEI (.sgml) o en formato RTF. El listado de obras de esta sección puede consultarse ordenado alfabéticamente por autor, cronológicamente o por géneros. Sin embargo, no todos los textos disponen de la misma cantidad de información. En el mejor de los casos encontraremos la siguiente: (1) El texto en sí, que, a diferencia de *Euskal Testuen Gordailua*, se presenta con la grafía modernizada; el problema es que los textos contienen muchos errores, ya que en ocasiones la modernización parece haberse realizado de forma (semi)automática. (2) En el margen derecho, se sitúan los enlaces a las secciones en las que se divide el texto, si este es de cierta extensión. (3) El nombre del autor enlaza con su posición en el listado alfabético. (4) El año, a su vez, nos sitúa la obra en el listado cronológico. (5) Enlace para descargar la obra completa en formato RTF. (6) Enlace para descargar la obra completa en formato PDF. (7) En algunos casos, enlace para descargar una edición facsimilar en formato PDF. (8) Se complementa, en el caso de los autores más estudiados, con otro servicio incluido en el macro-proyecto *Armiarma*, llamado *Literaturaren zubitegia* [Los puentes de la literatura], que incluye información diversa sobre la obra o el autor, bibliografía, traducciones, un directorio de enlaces a facsímiles de diferentes bibliotecas o a otras páginas web relacionadas, e incluso a textos en formato EPUB. (9) Fuente impresa utilizada y/o autor de la versión electrónica incluida.

La segunda sección lleva por título "Herri literatura" [Literatura popular] y está dividida en cuatro subsecciones: (1) "Baladak, poesiak, kantak" [Baladas, poesías, canciones], compuesta principalmente por baladas y poesías antiguas de los ss. XIV al XVII y algunas colecciones de canciones populares y *bertsos* del s. XIX. (2) "Ipuinak, kondairak" [Cuentos, leyendas], que enlaza con tres obras que aparecen también en la sección de autores y obras. (3) "Antzerkia" [Teatro], integrada por una colección de *astolasterrak*, un tipo de teatro popular en verso propio del País Vasco Continental. (4) "Atsotitzak eta beste" [Refranes y otras cosas], en la que se incluyen ejemplos de refranes pertenecientes a varias colecciones paremiológicas de los ss. XVI y XVII.

La tercera sección es la llamada genéricamente "Testuak" [Textos], dividida a su vez en dos subsecciones: (1) "Testu zaharrak" [Textos antiguos], que contiene

frases sueltas, oraciones, cartas y otros textos breves, tanto literarios como administrativos, de todos los períodos abarcados por el repositorio (es decir, hasta la Guerra Civil). (2) "Epigrafikoak" [Epigráficos], inscripciones de época romana que incluyen algún nombre en euskera, principalmente teónimos.

Estas últimas dos secciones presentan unas características especiales. Primeramente, como hemos señalado más arriba, estas secciones no están incluidas en el motor de búsqueda. En segundo lugar, aparte del texto, la única información complementaria que ofrecen es la fuente impresa ("Bertsio honen iturria") o el autor de la versión electrónica utilizada ("Bertsio informatiko honen egilea") y, en ocasiones, la fuente original ("Non"). En tercer lugar, los textos se presentan sin modernizar, respetando la grafía y la disposición de la fuente impresa u original utilizada. Por último, en cada subsección las obras se ordenan cronológicamente (excepto las obras epigráficas, todas de época romana, que aparecen dispuestas por orden alfabético).

3.4. Portal Andrés de Poza

El Portal Andrés de Poza¹⁷ es un proyecto desarrollado en la Universidad de Deusto, y ha ido creciendo paulatinamente desde 2007, siempre bajo la dirección de la doctora Carmen Isasi. Entre los proyectos aquí expuestos, este es sin duda el más ambicioso en la aplicación combinada de criterios filológicos y estándares de marcación, siguiendo las propuestas de la Text Encoding Initiative (TEI).

Se trata de un corpus reducido de textos múltiples, entendiendo como tales aquellos que permiten la presentación alineada de diferentes versiones o ediciones del mismo texto, o bien diferentes traducciones del mismo texto. Por tanto, el corpus no es exclusivamente (ni siquiera mayoritariamente) en lengua vasca, y de hecho los textos antiguos vascos incluidos son contados: *Platica lelengoá* de Rafael Mikoleta (1653), *San Frances Sales Genevaco Ipizpicuaren Philotea eta Chapeletaren Andre Dana Mariaren ohoretan devocionearequin erraiteco antcea* de Silvain Pouvreau (1664), *Gudu espirituala* de Silvain Pouvreau (1665) y *Philotea edo devocioneraco bide erakusçaillea* de Joannes Haraneder (1749).

Todos ellos comparten la particularidad de ser traducciones y son presentados en versión múltiple alineada, junto con sus originales respectivos o con traducciones de otras lenguas, que pueden ser cotejados fácilmente por pares, tríos y hasta cinco versiones diferentes. Los textos vascos han sido editados *ad hoc* para este proyecto por filólogos de prestigio (Josu Bijuesca y Ana Toledo), y los criterios de edición se explicitan detalladamente. Todos ellos cuentan con un aparato crítico cuidadoso (quizás no suficientemente audaz en la *emendatio*), aunque la presentación de variantes y notas paleográficas no es muy lograda desde el punto de vista visual. Según apuntan los responsables del proyecto, se optó por elaborar una edición conservadora de los testimonios, manteniendo en general la ortografía de los originales.

Además de la cabecera que recoge los datos fundamentales de cada texto y edición, han tratado mediante etiquetas TEI elementos como intervenciones de los editores, deterioros, citas o cambios de lengua, identificación y anotación de antropónimos y topónimos y correspondencias de segmentos.

¹⁷ <<http://andresdepoza.com/>>.

En todo caso, se trata de un proyecto en fase de experimentación que merece los apoyos necesarios para explorar nuevas posibilidades.

El portal ha optado por una licencia Creative Commons BY-NC-ND.

3.5. *Lazarragaren eskuizkribua: edizioa eta azterketa* [El manuscrito de Lazarraga: edición y estudio]

El sitio web¹⁸ que vamos a tratar en esta sección está dedicado a la edición y estudio de la obra de un único autor, pero tiene la particularidad de ser el primer texto antiguo vasco cuya edición filológica ha sido pensada para ser editada en internet como *work in progress*. Es, seguramente, a día de hoy el proyecto más ambicioso y completo desde el punto de vista de la filología vasca y se enmarca en el grupo de investigación *Monumenta Linguae Vasconum* de la UPV/EHU (Bilbao *et al.*, 2010).

La obra editada es el llamado *Manuscrito de Lazarraga*¹⁹, un cartapacio incompleto descubierto en 2004, compuesto por textos que pueden datarse, seguramente, entre la segunda mitad del siglo XVI y principios del XVII (entre 1567 y 1602, aproximadamente).

El sitio web *Lazarragaren eskuizkribua: edizioa eta azterketa* está publicado bajo una licencia Creative Commons BY-SA, lo cual no ha evitado que sea citado de manera deficiente por varios investigadores. Está estructurado en varias secciones, que comentaremos brevemente a continuación: (1) La introducción reseña la vida y obra del autor principal del manuscrito, Juan Pérez de Lazarraga, y expone los criterios seguidos en la edición. (2) La gramática se encuentra todavía en fase de elaboración; en estos momentos, el sitio web sólo dispone de un índice provisional de la futura gramática. (3) El diccionario también está en fase de elaboración; por ahora se ofrecen las concordancias del texto "en bruto", descargables en formato PDF. (4) La sección "Versiones" detalla el registro de los cambios realizados en el sitio web en cada nueva versión. (5) Una bibliografía recoge las ediciones previas del manuscrito y los estudios sobre este o sobre Lazarraga en general, así como todas las obras que se han utilizado y mencionado en el trabajo; cuando existe una versión en la red de alguno de los trabajos recogidos en la bibliografía, se añade el enlace correspondiente. (6) La "Ayuda" brinda algunos consejos sobre cómo utilizar el sitio web. (7) Finalmente, la sección central está dedicada a la edición del manuscrito, que se ofrece en distintas versiones, algunas más avanzadas que otras, y que pasamos a detallar.

La primera versión del texto a la que nos referiremos es la edición crítica, que se presenta tanto en HTML para leer directamente en la red, como en versión PDF para descargarse o imprimir. Esta primera versión se centra en la transcripción semipaleográfica del texto (con regularización –no modernización– de la grafía, actualización de la puntuación, del uso de mayúsculas y de la separación de palabras, entre otros cambios). A la transcripción de cada texto precede una introducción que explica brevemente su contenido, estructura, métrica, fuentes, etc. Por último, la edición cuenta con un aparato de notas paleográficas y otro

¹⁸ Accesible desde estas dos direcciones: <<http://www.ehu.es/monumenta/lazarraga/>> y <<http://lazarraga.com>>.

¹⁹ Para más detalles sobre el manuscrito, ver Gómez y Urgell (2010) y Gómez (2011).

de notas explicativas. Las notas paleográficas, en la versión HTML, se anuncian mediante números volados y están situadas al final de cada texto.

Por su parte, en esta presentación en HTML las notas explicativas están marcadas mediante hiperenlaces y situadas en la columna derecha de la página web. El objetivo de estas notas es ofrecer al lector todas las claves para poder entender el texto del modo más completo y correcto posible. Para decidir la cantidad, tamaño y profundidad de las notas, se ha tenido en mente a un lector potencial que ya posee un conocimiento mínimo de la historia del euskera y de sus dialectos.

Como hemos señalado, la edición crítica cuenta también con una presentación en formato PDF. Se trata de la misma edición que se presenta en HTML, pero con algunas modificaciones en cuanto a su disposición: las notas explicativas son aquí notas a pie de página y las notas paleográficas aparecen al final de cada texto.

La segunda versión corresponde al texto en euskera con grafía modernizada, presentado en formato PDF. En esta versión se han suprimido las notas paleográficas, y las notas explicativas se han reducido a las necesarias para la comprensión del texto.

La tercera versión es la traducción del texto al castellano (actualmente en fase de elaboración). Se presenta únicamente el texto, sin aparato crítico.

Por último, cada texto se enlaza con la reproducción facsimilar digitalizada que publicó la Diputación de Gipuzkoa en 2004, tras presentar el manuscrito en sociedad.

4. NUEVA RECOPIACIÓN DE TEXTOS ANTIGUOS VASCOS

El equipo *Monumenta Linguae Vasconum* de la UPV/EHU, dirigido por el catedrático Joseba A. Lakarra y mencionado en el apartado anterior, trabaja actualmente en una nueva recopilación de textos antiguos vascos, que provisionalmente denominamos con el nombre genérico correspondiente en lengua vasca: *Euskal Testu Zaharren Bilbuma Berria*.

Este trabajo pretende recoger el testigo de *Textos Arcaicos Vascos* de Michelena (1964) y de *Contribución al estudio y edición de textos antiguos vascos* de Ibon Sarasola (1983), concebido en su momento precisamente como continuación del anterior. El proyecto aspira a recopilar y editar todo texto²⁰, no necesariamente literario, escrito en lengua vasca antes de 1745, de menos de 14.000 caracteres; será a su vez complementario con una crestomatía de textos vascos más extensos del mismo período y parte de la experiencia de la edición y estudio de la obra de Lazarraga, expuesta en el apartado inmediatamente anterior a este.

El equipo de trabajo simultaneará la elaboración de una edición en papel y una edición *online*, complementarias una de la otra, pero concebidas desde su inicio pensando en las características del soporte respectivo. La edición en papel recogerá la edición filológica convencional (semipaleográfica) de los textos, y la edición *online* ofrecerá no sólo la edición filológica de los textos, sino también

²⁰ Quedan excluidas del concepto de texto las palabras sueltas que no llegan a formar al menos una frase, así como los topónimos, antropónimos, teónimos... que no formen parte de una frase redactada en lengua vasca.

la edición paleográfica y los textos en grafía modernizada, así como las reproducciones de originales y/o primeras ediciones, junto con las notas bibliográficas completas y, en su caso, las reproducciones o los enlaces de las principales ediciones modernas de cada uno de los textos.

En la edición *online*, la edición filológica de cada texto en formato HTML permitirá hacer uso de todas las virtualidades de este tipo de edición (hiperenlaces, multimedia...), pero se adjuntarán también los PDF de cada texto y de la obra completa, en sus diferentes versiones (edición filológica, paleográfica y modernizada).

Solamente la versión *online* acogerá, también por motivos obvios, el léxico de todo el corpus, que será presentado de dos formas: por una parte, las concordancias bruto, tal como son generadas por la herramienta informática, para que los investigadores puedan generar sus propias búsquedas y sus propios subproductos; por otra parte, un léxico elaborado por el equipo, con criterios unitarios para todo el corpus.

Finalmente, una base de datos exhaustiva permitirá habilitar un sistema de búsquedas con filtros de cronología, variedad dialectal, editor moderno... así como palabras, secuencias de palabras...

BIBLIOGRAFÍA

- Bilbao, Gidor, "Textos Arcaicos Vascos 50 urte geroago", en J. Gorrochategui, R. Gómez, J.A. Lakarra y C. Mounole (eds.), *Actas del congreso Koldo Mitxelena*, Bilbao, UPV/EHU, 2013, pp. 717-728.
- Bilbao, Gidor *et al.*, *Lazarraga eskuizkribua: edizioa eta azterketa (1.0)*, Vitoria-Gasteiz, UPV/EHU, 2010, <<http://lazarraga.com>> [24/09/2013].
- Euskal Testuen Gordailua*, Patxi Salaberri Muñoa (resp.), Vitoria-Gasteiz, UPV/EHU, 2000, <<http://www.vc.ehu.es/gordailua/>> [23/09/2013].
- Gómez, Ricardo, "Presencia e influencia del castellano en el manuscrito Lazarraga", *Oihenart*, 26 (2011), pp. 231-258.
- Gómez, Ricardo y Urgell, Blanca, "Hacia una edición crítica del manuscrito Lazarraga", *Oihenart*, 25 (2010), pp. 63-88.
- Gorrochategui, Joaquín; Igartua, Iván y Lakarra, Joseba A. (eds.), *Euskararen historia*, Vitoria-Gasteiz, Gobierno Vasco, 2013.
- Klasikoen Gordailua*, Zarautz, Susa, 2013, <<http://klasikoak.armiarma.com/>> [24/09/2013].
- Lakarra, Joseba A., "Euskararen historia eta filologia: arazo zahar, bide berri", *ASJU*, 31:2 (1997), pp. 447-535.
- Michelena, Luis, *Textos Arcaicos Vascos*, Madrid, Minotauro, 1964.
- OEH-ko testu corpusa* [CD-ROM], Bilbao, Euskaltzaindia/Aurten Bai Fundazioa, 2000.
- Sarasola, Ibon, *Historia social de la literatura vasca*, Madrid, Akal Bolsillo, 1976.
- Sarasola, Ibon, "Contribución al estudio y edición de textos antiguos vascos", *ASJU*, 17 (1983), pp. 69-212.



RESUMEN

En este artículo se presentan y evalúan los diferentes proyectos desarrollados hasta la actualidad para facilitar el acceso en internet a textos antiguos vascos, entendiendo el concepto de "texto antiguo vasco" como texto no necesariamente literario, de diversa extensión, escrito en lengua vasca antes de 1745.

Se presentan las principales bibliotecas digitales que permiten el acceso a facsímiles de ediciones impresas y manuscritos de textos antiguos vascos; se subraya la necesidad de un directorio exhaustivo que permita localizar fácilmente estos documentos.

A continuación, se ofrece una evaluación exhaustiva de las fortalezas, debilidades y virtualidades de cinco proyectos muy diferentes en su concepción, pero que tienen en común la presentación en formato digital de textos antiguos vascos: *Euskal Testuen Gordailua* [Depósito de textos vascos], *Corpus del Diccionario General Vasco – Orotariko Euskal Hiztegia* (OEH), *Klasikoen Gordailua* [Depósito de clásicos vascos], *Portal Andrés de Poza* y *Lazarragaren eskuizkribua: edizioa eta azterketa* [El manuscrito de Lazarraga: edición y estudio].

Finalmente, presentamos el proyecto en el que trabaja actualmente el equipo *Monumenta Linguae Vasconum*: una nueva recopilación de textos antiguos vascos, que pretende avanzar a partir de la experiencia parcial de la edición del manuscrito de Lazarraga.

Palabras clave: Textos antiguos vascos, internet, edición filológica, repositorios digitales.

ABSTRACT

This article presents and evaluates the different projects developed so far with the aim of facilitating access to old Basque texts on the Internet, the concept of "old Basque text" meaning a text not necessarily literary, of diverse length and written in Basque before 1745.

First of all, the article presents the different digital libraries that allow access to facsimiles of print editions and manuscripts of old Basque texts, highlighting the necessity of an exhaustive directory that will help find those documents easily.

Next, it offers a detailed evaluation of the strong and weak points as well as the potentialities of five projects which, although they were conceived very differently, are similar in that they all present old Basque texts in digital format: *Euskal Testuen Gordailua* [Depository of Basque texts], *Corpus del Diccionario General Vasco – Orotariko Euskal Hiztegia* (OEH) [Corpus of the Basque General Dictionary], *Klasikoen Gordailua* [Depository of Basque classics], *Portal Andrés de Poza* [Andrés de Poza web portal] and *Lazarragaren eskuizkribua: edizioa eta azterketa* [Lazarraga's manuscript: editing and study].

Finally, the article presents the project on which the *Monumenta Linguae Vasconum* research group is currently working: a new compilation of old Basque texts that aims to make progress from the partial experience of the editing of Lazarraga's manuscript.

Keywords: Old Basque texts, Internet, scholarly editing, digital repositories.

Indexación y códigos en bases de datos bibliográficas

JEAN-MARC BUIGUÈS
Université Michel de Montaigne Bordeaux3
JEAN-MARC.BUIGUES@U-BORDEAUX3.FR

INTRODUCCIÓN

En la primera parte de esta exposición presentaré a grandes rasgos el proyecto de bases de datos denominado NICANTO (Buiguès, 2011; Dedieu, 1997: 325-336). A continuación, me centraré en los problemas de indexación y codificación de datos en bases de investigación en las humanidades digitales, apoyándome en ejemplos concretos y en la experiencia acumulada en NICANTO. Sucesivamente evocaré la indexación y codificación de fechas, de lugares y de materias y la metodología seguida en el análisis de la sociología de los autores.

1. PRESENTACIÓN DE NICANTO

El proyecto NICANTO nació hace diecinueve años en la Maison des Pays Ibériques de la Universidad Michel de Montaigne Bordeaux3, un centro de estudios que funcionaba como una UMR (Unidad mixta de Investigación cuyo nombre era TEMIBER) que asociaba investigadores del CNRS y de la Universidad. Iniciado por tres investigadores: Jean-Pierre Dedieu (Director de investigaciones el CNRS) el añorado profesor François Lopez (Catedrático de la Universidad Michel de Montaigne Bordeaux3) y su servidor, en aquel momento recién nombrado Profesor Titular en dicha Universidad. Unimos nuestra experiencia en la investigación. François Lopez era el que mejor conocía la historia de la edición y de la imprenta españolas, también proyectaba informatizar la *Bibliografía de autores españoles del siglo XVIII* de Francisco Aguilar Piñal (Aguilar Piñal, 1981-2002) y los anuncios de impresos publicados en el siglo XVIII en la *Gaceta de Madrid*; Jean-Pierre Dedieu tenía una gran experiencia en bases de datos, experiencia acumulada en particular en la elaboración de la base FICHOZ sobre la alta administración española en la época moderna; yo llegaba con la experiencia adquirida durante la redacción de mi tesis doctoral (Buiguès, 1991) asentada en una base de datos sobre inventarios *post mortem*, testamentos, etc. y en la cual dedicaba una parte al análisis de los lectores y de las lecturas de los leoneses en el siglo XVIII a partir del estudio de sus bibliotecas –para elaborar un proyecto común bautizado “NICANTO” en homenaje a Nicolás Antonio.

El proyecto NICANTO se ideó como un sistema de bases de datos relacionadas entre ellas (ver Fig. 1). El núcleo central lo constituyó la primera base "Gnledici" en la que se informatizaron los ocho primeros tomos (los dedicados a obras con autoría) de la *Bibliografía de autores españoles del siglo XVIII*. Una segunda base, "Gnlrespo" que reunía a los autores fue creada a partir de Gnledici. La informatización de inventarios de bibliotecas privadas e institucionales españolas del siglo XVIII se empezó en dos otras bases. Una quinta base sirvió para identificar las fuentes utilizadas en las bases sobre inventarios de bibliotecas y reunir informaciones cualitativas sobre los poseedores de libros (calidad, oficios y cargos, niveles de fortuna, presencia de cuadros, etc.). La penúltima base fue la dedicada a la informatización de los anuncios de impresos publicados en el siglo XVIII en la *Gaceta de Madrid*, siendo la última base denominada "Censura" la que pretendía informatizar los índices inquisitoriales del siglo XVIII.

Actualmente la base Gnledici totaliza unas 26.000 fichas de 80 campos cada una, la base Gaceta, unas 19.000 fichas con 62 campos, las bases de inventarios de bibliotecas, unas 15.000 fichas con 60 campos, Gnlrespo, unas 7.600 fichas con 24 campos.

En realidad, el proyecto NICANTO se puso en marcha para poder desarrollar varios ejes de investigación:

- Historia de la edición y de la imprenta¹.
- Historia de los autores².
- Publicidad del impreso (anuncios de la *Gaceta de Madrid*)³.
- Lectores y lecturas, bibliotecas privadas y bibliotecas institucionales⁴.
- Evolución de la censura en el siglo XVIII.

Queda patente que a lo largo de estos casi 20 años de investigación, el proyecto inicial sufrió evoluciones debidas entre varios factores principalmente a la evolución de la estructura institucional de la investigación (Desaparición de la UMR TEMIBER), a las vivencias personales de los investigadores que constituyeron el núcleo fundador del proyecto, a las evoluciones técnicas en el mundo de la informática y de Internet, etc. Entre los ejemplos más significativos de esta evolución están la participación de NICANTO desde hace cuatro al proyecto PHEBO (Poesía Hispánica en el Bajo Barroco⁵) y más recientemente su participación a Dialogyca (Biblioteca Digital de Diálogo Hispánico⁶).

Sin embargo, las investigaciones llevadas a cabo permitieron enfrentarse con toda una serie de problemas que hoy en día siguen siendo fundamentales para quien desea basar su investigación en bases de datos relacionales como por ejemplo la homogeneización de los datos o los métodos que permitan acelerar

¹ Unos primeros resultados se publicaron en Infantes, López y Botrel (2003) y más precisamente en Buiguès (2003: 303-316) y Buiguès (2003: 317-327).

² El lector encontrará un ejemplo de la investigación sobre esta temática en Buiguès (2003: 292-302)

³ Un ejemplo de la investigación sobre este eje en Jean-Marc Buiguès (2013: 33-47).

⁴ Ejemplos de la investigación que corresponde a este eje en Infantes, López y Botrel (2003: 424-433) y Buiguès (2008: 361-367).

⁵ Ver PHEBO, <<http://phebo.es/es/presentacion>>.

⁶ Ver Dialogyca BDDH (Biblioteca Digital de Diálogo Hispánico), <http://pendientedemigracion.ucm.es/info/dialogycabddh/f_presentacion.html>.

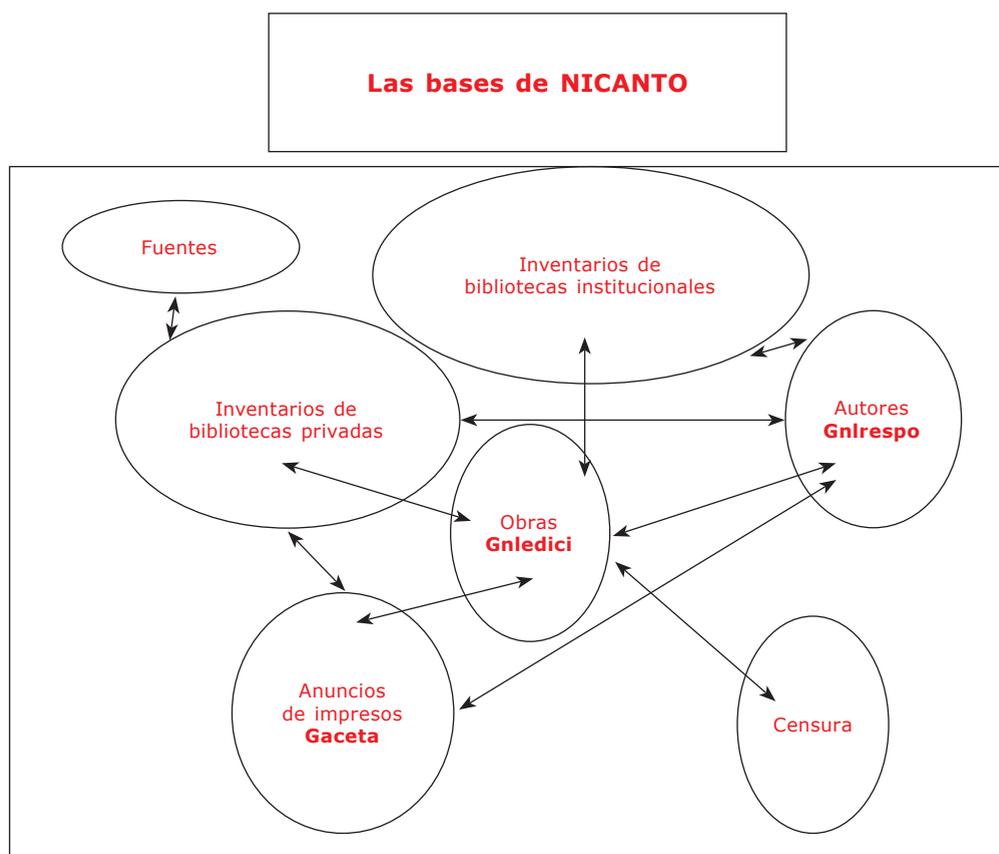


Fig. 1.

las búsquedas. Entre los procedimientos que se utilizaron para resolver dichos problemas presentaré a dos de ellos: la indexación y la codificación.

2. INDEXACIÓN Y CODIFICACIÓN DE DATOS EN NICANTO

Antes de presentar algunos ejemplos de indexación y de codificación⁷ quizás sea necesario dar unas cuantas definiciones. Indexar es en informática, la "acción de ordenar registros empleando índices. Un ejemplo son las bases de datos, que emplean uno o más índices para identificar registros", siendo el índice en una base de datos "una o más claves que identifican un único registro (fila o renglón). Los índices hacen más fácil encontrar y ordenar los registros"⁸. El término codificación es el "Proceso de organización de la información en

⁷ Es de señalar que "indexación" y "codificación" no figuran todavía con estos significados en el DRAE.

⁸ Ambas definiciones provienen del *Diccionario de informática*, <<http://www.alegsa.com.ar/Dic/indexar.php>>.

categorías, mediante el cual se asignan códigos para introducir, almacenar y recuperar datos”⁹.

Estos procesos no son sino procesos de transformación de la información de la fuente primaria, o dicho de otro modo procesos que participan en distintos momentos de la investigación que son la entrada, el almacenamiento y la búsqueda de la información.

2.1. Ejemplos de indexación en NICANTO

Un primer ejemplo de indexación es el que ofrece el campo “Nº_Edición” en la base Gnedici. Este campo es alfanumérico. Empieza por letras seguidas por dígitos. Veamos unos cuantos ejemplos: “BEGUE1”, “AP-A01396”¹⁰. En el caso de “AP-A01396”, AP indica que la fuente de la ficha es la *Bibliografía de autores españoles del siglo XVIII* de Francisco Aguilar Piñal; la letra “-A” (“AP-A01396”), el tomo 1 y “01396” el número de identificación de la ficha en la *Bibliografía*. Para “BEGUE1”, “BEGUE” indica que la fuente es la “Relación de la poesía española publicada entre 1648 y 1750” (Bègue, 2010) y “1” que es la primera obra mencionada en dicha relación. Se hubiera podido utilizar para este campo “Nº_Edición” únicamente un número pero el sistema de indexación elegido permite buscar en función de la fuente. Por ejemplo una búsqueda en dicho campo con “AP-B” me da todas las fichas del segundo tomo de la *Bibliografía* lo que una indexación meramente numérica no hubiera permitido.

Del mismo modo en la base *Gaceta*, el campo “Número” se rellena de la manera siguiente “GM-004093” en que “GM” significa la *Gaceta de Madrid* y “004093” el número de la ficha por orden de entrada de datos. La mención de “GM” permite prever la futura y/o posible informatización de anuncios de impresos de otros periódicos. Por ejemplo si se tratase de anuncios del *Memorial literario* se podría rellenar el campo con “ML”.

2.2. Codificación de fechas

Para establecer la cronología de la edición en la España del siglo XVIII fue necesario profundizar la utilización de las fechas tal como se mencionan en las fuentes primarias. Las fechas presentes en las fuentes primarias ofrecen según las fuentes una gran variedad desde su total ausencia como en el caso de los romances populares publicados en el siglo XVIII hasta la pormenorizada mención del año, mes y día como lo ofrecen los anuncios de impresos de la *Gaceta*. Si examinamos por ejemplo las fechas de publicación en Gnedici uno de los problemas que se plantean es por ejemplo el de las obras con varios volúmenes publicadas durante varios años. La dieciséis volúmenes de la “Synopsis histórica cronológica de España. parte primera que comprende sus sucesos desde la creación de el mundo hasta el nacimiento de Jesu Christo ...: formada de los autores seguros y de buena fee / por don Juan de Ferreras ...impresa en Madrid por Francisco de Villa-Diego: a costa de Diego Lucas

⁹ Según definición de *Los diccionarios y las enciclopedias sobre el Académico*, <http://www.esacademic.com/dic.nsf/es_mediclopedia/30490/>.

¹⁰ Al proyecto inicial que sólo preveía la informatización de la *Bibliografía* de Aguilar Piñal se han añadido otros datos procedentes de varias fuentes: bibliografías impresas, recursos digitales, etcétera.

Codificación cronológica

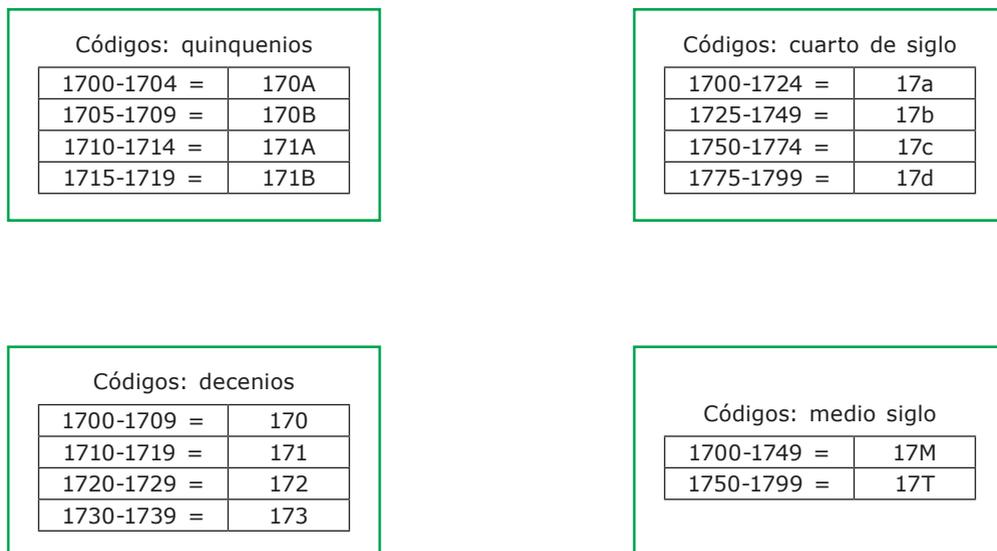


Fig. 2.

Ximenez¹¹ se publicaron entre 1702 y 1727. El campo "Fecha_o" se rellenó así: "1702-1724". Cuando el investigador hace una búsqueda por fechas en el campo "Fecha_o" sólo saldrá la ficha si tecllea "1702" o "1727". Es menester entonces crear campos de codificación de las fechas. La figura 2 indica cuatro modalidades de codificación de las fechas: por quinquenios, por decenios, por cuartos de siglo o por mitades de siglo¹². En el ejemplo de la "Synopsis" de Juan de Ferreras se rellenará el campo "medio siglo" con "17M" (ver Fig. 2). En función de la precisión de la fecha de edición –cuando la obra carece de fecha de publicación las fechas que servirán para rellenar el campo ""Fecha_o" pueden ser por ejemplo las que corresponden al período de actividad del impresor, y por consiguiente los campos correspondientes de codificación de las fechas, etc.– se rellenarán todos los campos o solamente algunos. La codificación de las fechas permite por ejemplo trazar la curva de la producción de obras nuevas en castellano durante el siglo XVIII como lo evidencia el gráfico de (Fig. 3) (Infantes, López y Botrel, 2003). En este caso por la abundancia de datos fue posible utilizar los

¹¹ Datos que provienen de la *Bibliografía* (tomo 3, nº 3781). Se respeta la ortografía de la fuente.

¹² Se crearán cuatro campos: "Fecha_quinquenio", "Fecha_decenio", "Fecha- cuarto" y "Fecha_mitad".

**Producción de obras nuevas en castellano
(curva anual y quinquenal)**

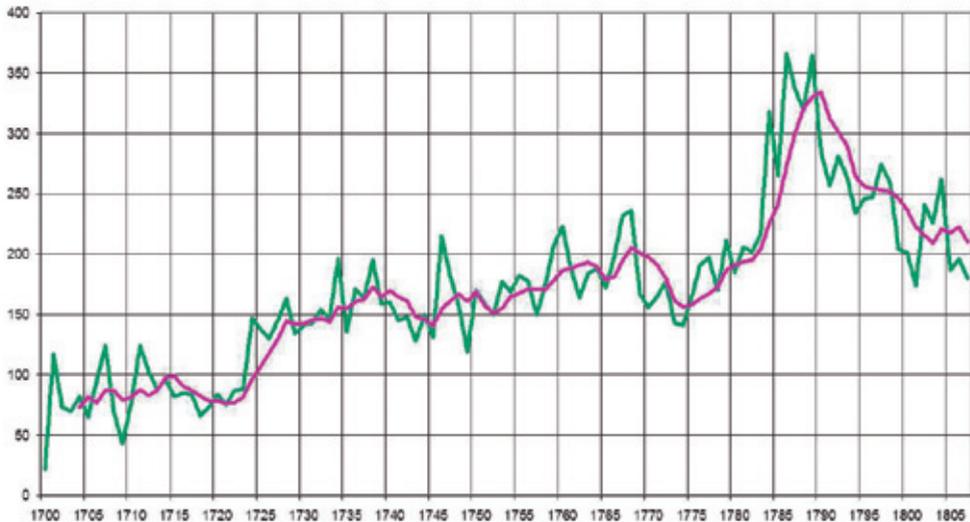


Fig. 3.

campos "Fecha_o" y "Fecha_quinquenio" porque había datos suficientes como para establecer las dos curvas. En el caso por ejemplo del romancero popular del siglo XVIII, la ausencia muy importante de la mención del año de edición sólo permite y ni siquiera siempre rellenar el campo "Fecha-mitad" para medir las evoluciones del género sólo entre la primera y la segunda mitad del siglo.

2.3. Codificación de lugares

En la base Gnedici se creó un campo "Lugar_O" para el lugar de edición de la obra tal como aparece en la fuente. Sin embargo no sólo las variantes ortográficas (B/V, etc.) sino también la presencia de los nombres de las ciudades en latín y en castellano (pero también a veces en otros idiomas), dificulta las búsquedas. Por ejemplo para la ciudad de Lyon, el investigador tiene que buscar sucesivamente por "Lyon", "León" y Lugduni". Además en el caso de la búsqueda por "León" aparecerán no sólo fichas de "León de Francia" según la terminología de la época sino también las de la castellana ciudad de León y las de la Isla de León, ciudad ésta que por otra parte aparece también en otras fichas designada por San Fernando. Para homogeneizar estos datos, evitar confusiones y facilitar las búsquedas, se creó un campo "Lugar_H" donde se pone el nombre moderno: "Lyon", "León" para la ciudad castellana y "San Fernando" para la Isla de León. Así el investigador que busque por "Lyon" podrá localizar en una sola búsqueda todas las fichas de Lyon, León y Lugduni.

Por otra parte, unos de los ejes de la investigación era establecer la geografía de la edición lo que necesitó la creación de un campo de codificación de esta información. Por ejemplo para medir la producción del impreso poético español publicado en Europa entre 1650 y 1699, hacía falta un campo de codificación

que permita distinguir la producción francesa de la italiana por ejemplo. Se creó para dicho fin el campo "Cod_Lugar" con tres dígitos. Por ejemplo en el caso de "EFA" la letra "E" corresponde a Europa, la letra "F" a Francia y la "A" designa París. Se puede con este campo saber pues cuántas obras de poesía española se publicaron en Europa, en Francia o en París. Esta codificación de los lugares permitió establecer el mapa de las ediciones de impresos poéticos españoles en Europa (Buiguès, 2013) y (López, Sáez y Sánchez, en prensa) (ver Fig. 4).

El campo "Cod_Lugar" permite también llevar a cabo estudios regionales como por ejemplo medir la producción gallega en el siglo XVIII. En el cuadro siguiente la letra "I" corresponde a Galicia, mientras las letras "A", "B", "C" y "D" a las ciudades de respectivamente La Coruña, Santiago de Compostela, El Ferrol y Orense.

	<i>Ediciones</i>	<i>%</i>	<i>Cod_Lugar</i>
Orense	2	1,0	PID
El Ferrol	2	1,0	PIC
La Coruña	23	11,2	PIA
Santiago de Compostela	179	86,9	PIB
Total Galicia	206		

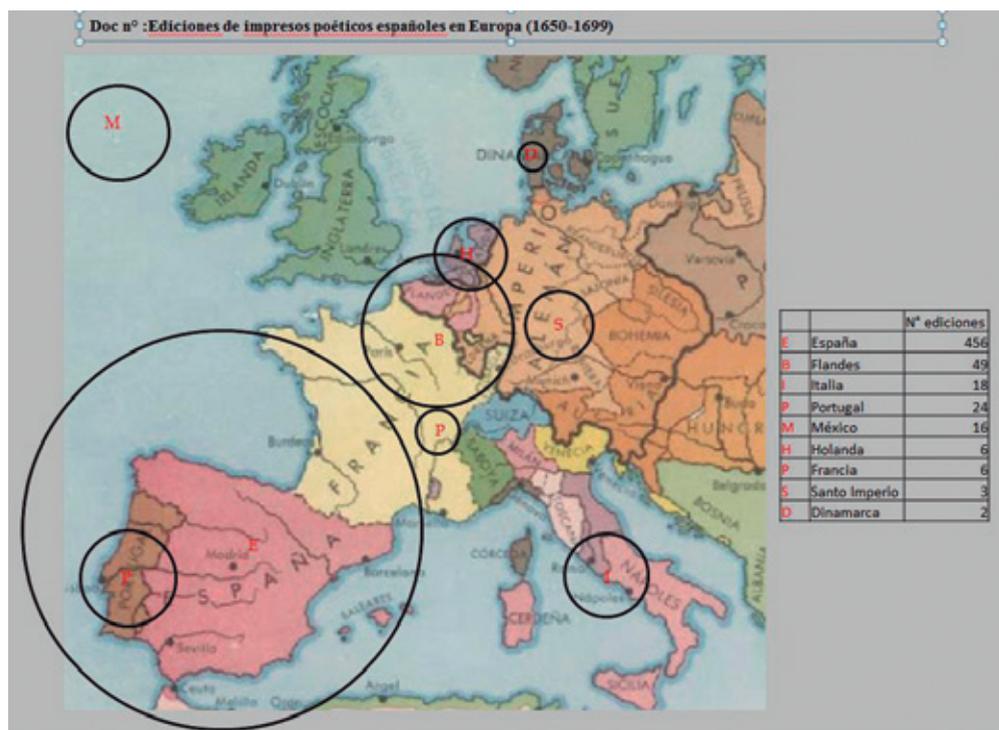


Fig. 4.

Otra base importante de NICANTO es la base que reúne los anuncios de impresos de la *Gaceta de Madrid* entre 1697 y 1807. Entre los datos que ofrecen los anuncios figura el lugar de venta del impreso que puede ser un local de profesionales (libreros, impresores) pero también la casa del autor, la portería de un convento, una tienda de comestibles, etc. Además de la variedad tipológica de los puestos de venta (librero, portería de convento, etc.) un mismo puesto puede mencionarse bajo distintas formas. Por ejemplo, los propios anuncios de la *Gaceta* se anuncian a veces "Donde la Gazeta" y otras "Donde estas relaciones se hallará". Otro elemento importante es que los anuncios no son sólo de Madrid, los hay también de otras ciudades. Fue necesario crear un campo de codificación y homogeneización de estos datos: el campo "CodPuestoVenta". A continuación se dan algunos ejemplos de esta codificación:

Md180 = Puesto de Francisco Robledo, Puerta de Guadalajara
 Md181 = En la imprenta de Joseph Rodriguez, Calle del Carmen
 Vd3 = Casa de la viuda de Santánder
 VI8 = Casa de Manuel Cabero, calle de Campaneros

Las letras "Md", "Vd" y "VI" corresponden a Madrid, Valladolid y Valencia, los números a un lugar de venta. Este campo permite medir la importancia de cada ciudad en cuanto a la publicidad del impreso, pero también de ver para cada ciudad quiénes son los que anuncian en la *Gaceta*. También se creó un campo de codificación del tipo de lugar de puesto de venta llamado "CodTipoPuesto" en que por ejemplo "P" corresponde a un anuncio de portería de convento, "I" al de una imprenta, "L" al de un librero, "T" al de una tienda, etc. Utilizando este campo se puede por ejemplo para Madrid, evaluar el peso de los anuncios de profesionales del libro frente al de los no profesionales.

2.4. Sociología de autores

La base sobre autores, Gnlrespo, proporcionaba la posibilidad de hacer un estudio de la sociología de los autores dieciochescos. Para este fin se crearon varios campos de codificación de los cuales sólo mencionaré tres. "CATEG1" permite distinguir el estamento (nobleza, clero o estado llano), pero también por ejemplo para los eclesiásticos si pertenecen al clero regular o secular, de qué orden son o qué función o cargo desempeñan. Por ejemplo si en el campo CATEG1 tecleo "R-SJ" aparecen todas las obras cuyos autores son jesuitas o sea 373 autores que publicaron 7593 obras. Del mismo modo el campo CATEG2 permite hacer búsquedas a partir de los oficios de los autores: así "F-AA" permitirá localizar a los miembros de los Consejos Reales, "F-MI" a los ingenieros de marina y "M-AP" a los pintores. Sin embargo una búsqueda con sólo "M" hubiese permitido seleccionar a los autores que ejercitaban oficios mecánicos y artísticos, "M-A" a únicamente los artistas. El campo CATEG3 es el que permite hacer búsquedas para los miembros de Academias con el mismo sistema de informatización jerarquizada en función del rango de cada letra: por ejemplo si busco por "A-HV", la "A" corresponde a un miembro de una academia, la "H" indica que se trata de la Real Academia de la Historia y la "V" que el autor desempeña el cargo de director.

2.5. Codificación de Materias

Cuando se creó la base Gnlledi uno de los ejes de investigación previstos era el estudio de la producción impresa en función de las materias. Una solución

<p>Listado de materias</p>	<p>Academias, Antologías, Artículos periodísticos, Atribuciones, Autobiografías, Bibliografías, Biografías, Cancioneros, Cartas, Compilaciones, Crónicas, Documentos legales, Estudios, Facsimiles, Genealogías, Glosas, Historias de escritoras, Justas, Manuscritos, Menciones coetáneas, México, Obras dramáticas, Obras generales, Obras perdidas, Páginas web, Partidas de bautismo, Pliegos sueltos, Poemas, Portugal, Preliminares, Profesiones religiosas, Prosa de Ficción, Prosa de pensamiento, Prosa religiosa, Relaciones, Siglo XV, Siglo XV, Siglo XVI, Siglo XVII, Siglo XVIII, Tesis doctorales, Testamentos, Traducciones, Tratados</p>	7. TABLA DE TÉRMINOS ADMITIDOS DE "MATERIAS Y TEMAS SECUNDARIOS"		
<p>Otras materias de búsqueda</p>	<p>Nombres de autoras Siglo en que escribieron (números romanos) Género de las obras Estudios o ediciones</p>	<p>Adivinación Agricultura Alimentación Alquimia América Amos Asatruas Antropología Arbuzismo Aritmética Armas y letras Arqueología Arquitectura Arte Artes de bien morir Artes de gobernar Artes de la guerra Artes de navegar Artilería Ascética Astrología Astronomía Biología Botánica Catequesis Caza Cirugía Ciudadad Comercio Controversia religiosa Cosmografía Cosmología Cristianismo Crítica de costumbres Crítica social Danza Derecho Dialéctica Dietética Economía Edades de la vida Educación</p>	<p>Emblematiza Enseñanza de lenguas Escatología Escritura Españolidad Estados civiles Estramentos Estética Ética Enseña Emografía Farmacia Filografía Filosofía Filosofía moral Filosofía natural Finanzas Física Geografía Geología Geometría Gramática Guerra Heráldica Heresía Historia Honos Iconografía Ingeniería Islam Judaísmo Juegos Lenguas Lugares Literarias Lógica Locuaz Magia Máquinas Matemáticas Mecánica</p>	<p>Medicina Melancolía Metafísica Metaburgia Mística Mitología Mugs Música Naturaleza Navegación Negocios Numismática Odontología Oficios Óptica Órdenes militares Órdenes religiosos Ortografía Países Profesiones Poesía Punera Póetica Política Pucología Química Religión Retórica Sucesos particulares Superstición Teatro Teología Teoría de la historia Teoría de los sabores Territorios Textos sagrados Traducción Varia lección Veterinaria Vida de corte y de aldea</p>
<p>Siglo de una autora</p>	<p>De forma general se consigna el siglo en el que vive y escribe, no el de publicación de su obra. Autoras que viven entre dos siglos, se incluyen los dos</p>			
<p>Búsqueda combinada de materias</p>	<p>Se pueden combinar varios conceptos Se pueden usar ventanas diferentes para cada concepto o ponerlos en la misma ventana sin separación entre ellos</p>			
<p>Menciones coetáneas BIESES</p>	<p>Se refiere a citas y menciones de las autoras, hechas en su propio tiempo o poco después</p>			
<p>BIESES, BIBLIOGRAFÍA DE ESCRITORAS ESPAÑOLAS, pantalla de ayuda para realizar búsquedas en el campo materias, http://62.204.193.244:8080/bieses/ayuda/bieses.html#busquedapormaterias</p>				
<p>DIALOGYCA BDDH, BIBLIOTECA DIGITAL DE DIÁLOGO HISPÁNICO, Manual de procedimiento de trabajo normalizado, (3ª versión febrero 2009), p.125</p>				

Fig. 5.

posible hubiera sido crear un campo "Materia" y establecer un listado de las palabras que pudieses servir para rellenar dicho campo. Es por ejemplo el sistema que suelen utilizar las bases de datos bibliográficas como la de Biblioteca Nacional de España o la del Catálogo Colectivo del Patrimonio Bibliográfico Español. En regla general las búsquedas por materias en este tipo de bases no suelen dar resultados muy completos por distintas razones: por ejemplo en el Catálogo Colectivo, este campo se rellenó según criterios diferentes en función de la institución que conserva el ejemplar de una obra (más de setecientas instituciones) pero también con criterios que cambiaron a lo largo del tiempo para una misma institución. En el caso de bases especializadas elaboradas por grupos de investigación cada base elabora en función de sus ejes de trabajo soluciones distintas como se puede ver en las dos magníficas bases Dialogyca y Bieses (ver Fig. 5). La solución en Gnedici fue distinta: se crearon varios campos para las materias y formas. En un campo llamado MAT1 se establecieron, a partir de una clasificación de las materias según la clasificación elaborada para las bibliotecas del Antiguo Régimen por el historiador François Furet (Furet, 1965: 3-32)¹³, cinco grandes categorías. "A" = Religión, "B" = Derecho, "C" = Historia, "D" = Ciencias y Artes, "E" = Bellas Letras. Cada categoría se subdividía. Por ejemplo la categoría "Ciencias y Artes" se componía, siempre basándose en la clasificación

<p><u>D2.—ARTES</u></p> <p><u>B: Artes mecánicas</u></p> <p>1: Relojes</p> <p>2: Madera</p> <p>3: Seda</p> <p>4: Platería</p> <p>5: Hierro</p> <p>6: Cobre</p> <p>7: Lana</p> <p>8: Pirotecnia</p> <p>9: Fuego</p> <p>10: Vidrio</p> <p>11: Papel</p> <p>12: Cuero</p>	<p><u>D2.—ARTES</u></p> <p><u>A: Artes liberales</u></p> <p>1: Música</p> <p>2: Pintura</p> <p>3: Escultura</p> <p>4: Dibujo</p> <p>5: Grabado</p> <p>6: Arquitectura</p> <p>7: Arte militar</p> <p>8: Arte de escribir</p> <p>9: Decoración</p> <p>11: Danza</p> <p>12: Ópera</p>
--	--

Fig. 6.

de François Furet, de los bloques siguientes: D1.—Ciencias, D2.—Artes, D3.—Agricultura y agronomía, D4.—Juegos de destreza y pasatiempos, D5.—Economía y política, D6.—Diversos. Del mismo modo cada una de estas subcategorías se dividía en otros bloques. Por ejemplo "D2" que corresponde a Artes se subdividía a su vez en "D1B" para las artes mecánicas y en "D2A" para las artes liberales. Por fin estas dos subcategorías daban lugar a una última caracterización con el añadido de cifras: por ejemplo, "D2A8" identificaba las obras que versaban sobre el arte de escribir (ver Fig. 6). Este sistema de codificación permitía medir por ejemplo el peso de la producción de obras científicas en el conjunto de la producción pero también según lo que el investigador buscaba localizar por ejemplo a todas las obras que trataban de cirugía.

Sin embargo, como se mencionaba en la primera parte de este trabajo, las líneas de investigación fueron cambiando con el transcurso de los años lo que hizo que nuevos campos para las materias tuvieron que ser creados. Dentro de las últimas líneas de trabajo está la dedicada a la poesía posbarroca dentro del equipo que dirige Pedro Ruiz Pérez¹⁴. Para estudiar este tipo de impreso se han creado campos específicos. Por ejemplo un campo "Protagonista" en que se pone la identidad del sujeto sobre que versa el poema: "Santo Tomás de Aquino", "Carlos III", "doña Rosa de Guzmán Silva, dignísima hermana del duque de Medina Sidonia", etc.; un campo "Acontecimiento" en que se especifica por ejemplo "Nobleza, bodas", "Academia poética, certamen", "polémica teatral" y otros que no se mencionan aquí. En esta primera fase del trabajo se reelaboran tanto informaciones que da la fuente primaria como informaciones añadidas por el investigador. Sin embargo para tener una visión de conjunto será indispensable crear campos de codificación para por ejemplo los acontecimientos. Si se quiere analizar los poemas que tratan de episodios de la vida de la familia real y los de la nobleza es posible crear un campo "Cod_Acontecimiento" con un sistema jerarquizado de la información que podría ser el siguiente: la primera

¹⁴ "Poesía del período posbarroco (1650-1750): repertorios y categorías" (FFI2008-03415-E/FILO del MICINN).

letra indicaría si el protagonista pertenece a la familia real, etc. la segunda el tipo de acontecimiento, boda, nacimiento, defunción, etc. Así "RB" indicaría que se trata de las bodas de un miembro de la familia real, "NN, del nacimiento de un noble, etc. De este modo el investigador ahorraría tiempo y encontraría las debidas respuestas a las preguntas formuladas en su investigación.

CONCLUSIÓN

El primer propósito de este trabajo fue exponer el camino seguido en Nicanto para ir de la constitución de una base que hubiera podido ser meramente bibliográfica hasta la elaboración de una base de investigación. El elenco de ejemplos de indexación y de codificación presentados aquí demuestra que sin ellos hubiera sido casi imposible llegar a los resultados publicados. Por ejemplo por mucho que uno manejara las 23.000 fichas de la *Bibliografía* de F. Aguilar Piñal, le hubiera sido casi imposible establecer una curva de la producción del impreso español en el siglo XVIII o saber cuántas obras fueron publicadas por los jesuitas. La informatización de la monumental cantidad de datos recogidos por F. Aguilar Piñal, su indexación y codificación sí que lo permitieron.

Exponer el camino recorrido, señalar unas cuantas técnicas de homogeneización y codificación de datos apuntando cómo en definitiva fueron los distintos ejes de investigación y los propios investigadores los que definieron estos procesos era lo que se pretendía aquí.

Mas no sólo: me pareció también importante reflexionar o solicitar una reflexión sobre la distinción entre lo que son las bases de datos bibliográficos de consulta pública que no fueron, ni son todavía ni tampoco tienen por qué serlo, elaboradas para la investigación. Estas bases cuya grandísima utilidad no hace falta subrayar proporcionan una información primaria que viene a completar la información más tradicional que da el impreso o el manuscrito: son unas de las fuentes que permiten la creación de un segundo tipo de bases que elabora la investigación. En el caso de las bases de investigación como Nicanto, los investigadores van a añadir toda una serie de informaciones a las que proporcionan las fuentes primarias. Entre estas informaciones añadidas figuran como acabamos de verlo no sólo las que provienen de la necesidad de homogeneizar datos multiformes sino también los procesos de codificación de la información que permiten, sobre todo si la base cuenta con un gran número de fichas, contestar de manera rápida y eficaz a las problemáticas definidas por cada eje de investigación.

El vaivén constante entre la información que proporcionan las fuentes primarias y las que establece la investigación con la también necesaria posterior codificación son en definitiva lo que distingue una base de consulta abierta a todos y una base de investigación. Lo que no significa que una base elaborada para la investigación no pueda también pasar a ser base de consulta pública. El problema es la relación entre ambas o dicho de otro modo cuál es la información de la base de investigación que se da a ver libremente al público. La respuesta sea quizás compleja. Si algunas bases de investigación pueden dejar libre acceso a la totalidad de sus fichas es sin duda porque no permiten búsquedas elaboradas en función de cada nuevo investigador que consulte la base dado que suelen llevar pocos campos y ninguno codificado, y sobre todo porque

en los distintos campos la información no se ofrece de manera jerarquizada. Ciertos tipos de búsquedas como las evocadas anteriormente seguirían sin embargo siendo siempre posibles en estas bases pero no serían nada fáciles y llevarían mucho tiempo. Por otra parte poner a libre disposición todo cuanto fue elaborado por un investigador para su investigación tampoco es deseable porque permitiría a otros investigadores utilizar estas codificaciones para sus propias investigaciones sin que sea siempre posible comprobar la fuente. Esta información añadida en las bases de datos de investigación a través de los procesos de codificación es finalmente una nueva modalidad de la creación, y como tal forma parte del derecho a la propiedad intelectual. La solución consistiría sin duda a definir perfiles de usuarios que diesen acceso a distintos niveles de información.

BIBLIOGRAFÍA

- Aguilar Piñal, Francisco, *Bibliografía de autores españoles del siglo XVIII*, Madrid, CSIC, 1981-2002, 10 tomos.
- Bègue, Alain, "Relación de la poesía española publicada entre 1648 y 1750", en Aurora Egido y José Enrique Laplana (eds.), *La luz de la razón. Literatura y Cultura del siglo XVIII. A la memoria de Ernest Lluch*, Zaragoza, Institución "Fernando el Católico" (Col. Actas), 2010, pp. 399-477.
- Buiguès, Jean-Marc, *La ville de Léon et sa juridiction. Economie, société et mentalités en Vieille Castille au XVIII^{ème} siècle*, Tesis dirigida por Jacques Soubeyroux, Universidad de Montpellier, 1991.
- Buiguès, Jean-Marc, "Evolución global de la producción", en Víctor Infantes, François López y Jean-François Botrel (coords.), *Historia de la edición y de la lectura en España 1472-1914*, Madrid, Fundación Germán Sánchez Ruipérez, 2003, pp. 303-316.
- Buiguès, Jean-Marc, "Las materias: tradición y modernización", en Víctor Infantes, François López y Jean-François Botrel (coords.), *Historia de la edición y de la lectura en España 1472-1914*, Madrid, Fundación Germán Sánchez Ruipérez, 2003, pp. 317-327.
- Buiguès, Jean-Marc, "Géographie de l'édition des livres de l'abbé de la collégiale de Villafraanca del Bierzo (1795)", en Philippe Meunier y Edgard Samper (coords.), *Mélanges en hommage à Jacques Soubeyroux*, Saint-Etienne, Celec, 2008, pp. 361-367.
- Buiguès, Jean-Marc, "NICANTO. Obras impresas de escritores del siglo XVIII", en María Casas del Álamo, Germán Redondo Pérez y Sara Sánchez Bellido (eds.), *Actas del Seminario Internacional sobre Bibliotecas Digitales y Bases de datos Especializadas para la investigación en Literaturas Hispánicas* (BIDESLITE), Madrid, IUMP (UCM), 2013, <http://eprints.ucm.es/21207/1/Actas_Seminario_BIDESLITE.pdf>.
- Buiguès, Jean-Marc, "Les annonces d'imprimés littéraires dans la presse espagnole: 1697-1737", en Florence Boulerie (ed.), *La médiatisation du littéraire dans l'Europe des XVII^e et XVIII^e siècles*, Biblio 17, vol. 205, suplemento a Papers on French Seventeenth Century Litteratue, Tübingen, Narr/Verlag, 2013, pp. 33-47.
- Buiguès, Jean-Marc, "El impreso poético posbarroco: centro y periferia, impresores y formas materiales", *Versants*, nº 60:3 (fascículo español) (2013).
- Dedieu, Jean-Pierre, "El sistema Nicanto", *Bulletin Hispanique*, 99-1 (1997), pp. 325-336.
- Furet, François, "La 'librairie' du royaume de France au XVIII^e siècle", *Livre et société dans la France du XVIII^e siècle*, Paris, Mouton, 1965, pp. 3-32.
- Infantes, Víctor; Lopez, François y Botrel, Jean-François (coords.), *Historia de la edición y de la lectura en España 1472-1914*, Madrid, Fundación Germán Sánchez Ruipérez, 2003.



RESUMEN

El presente artículo presenta en su primera parte las bases de datos relacionales que constituyen el proyecto NICANTO del equipo de investigación AMERIBER de la Universidad de Burdeos III así como su evolución desde su creación hace veinte años. El proyecto abarca un amplio abanico de líneas de investigación: historia de la edición y de la imprenta, sociología de los autores, publicidad del impreso en la prensa, lectores y lecturas, bibliotecas privadas y bibliotecas institucionales, evolución de la censura en el siglo XVIII, poesía posbarroca (1650-1750), diálogos del siglo XVIII. El marco cronológico lo constituyen los siglos XVIII y XIX, el marco espacial España y Cuba. Las distintas bases contabilizan unas 75.000 fichas.

La segunda parte del artículo expone una serie de métodos elaborados en función de las distintas líneas de investigación. Estos métodos se centran en las nociones de indexación y de codificación de los datos informatizados a partir de fuentes primarias. Los ejemplos desarrollados corresponden tanto a indexación de la localización de las fuentes primarias como a codificaciones de campos de fechas, de lugares, de autores o de materias.

Por fin se sugiere una reflexión en torno a la distinción entre base de datos de consulta pública y base de datos para investigación.

Palabras claves: Bases de datos relaciones, España, siglo XVIII, edición, bibliotecas, codificación de datos.

ABSTRACT

This article presents in its first part the relational databases that constitute the project NICANTO of the AMERIBER research of the University of Bordeaux III team as well as its evolution since its creation twenty years ago. The project covers a wide range of research interests: history of editing and printing, sociology of the authors, advertising of the printed in the press, readers and readings, private libraries and institutional libraries, evolution of censorship in the 18th century, poetry posbarroca (1650-1750), dialogues of the 18th century. The chronological framework is the 18th and 19th centuries, the spatial context Spain and Cuba. The different foundations accounted for a 75,000 tokens.

The second part of the article exposes a number of methods elaborated according to the different lines of research. These methods focus on notions of indexing and encoding of the data computer from primary sources. Developed examples correspond to index the location of primary sources and encodings of fields of dates, places, authors or subjects.

Finally it is suggested that a reflection on the distinction between inspection database and database for research.

Keywords: Database relations, Spain, 18th century, Edition, libraries, data encryption.

El proceso de desarrollo de una biblioteca digital y su apertura e integración en el recolector *Hispana*: el caso de la *Biblioteca Digital Ovidiana Abierta*¹

PALOMA CENTENERA CENTENERA
Universidad Pontificia de Salamanca en Madrid
paloma.centenera@upsam.es

MARÍA LUISA DÍEZ PLATAS
Universidad Pontificia de Salamanca en Madrid
mluisa.diez@upsam.es

LA BIBLIOTECA DIGITAL OVIDIANA

Unos requisitos basados en la calidad

Desde un punto de vista técnico la biblioteca ha sido diseñada para ser una herramienta eficaz para usuarios e investigadores. Para ello, además de proporcionar un marco estético que acompañe la belleza de las ilustraciones de Ovidio que en ella se muestran, debe permitir a los usuarios encontrar fácilmente la información que necesitan.

Una biblioteca digital, como cualquier aplicación web, tiene unas características propias que diferencian su desarrollo del de otros sistemas informáticos. Estas características diferenciadas hacen que los atributos con los que se mide su calidad tengan también singularidades. Olsina *et al.* (Olsina, 1999) identifican los siguientes requerimientos de la calidad en aplicaciones web:

- Facilidad de uso. Coloquialmente este término hace referencia a la facilidad para utilizar una herramienta. La Oficina Internacional para la estandarización (ISO), en una referencia muy adecuada para nuestro fin, define este atributo como “la efectividad, eficiencia y satisfacción con la que un producto permite alcanzar objetivos específicos a usuarios específicos en un contexto de uso específico” (ISO/IEC 9241-11).
- Funcionalidad. Esta característica está relacionada con el valor añadido de los servicios proporcionados por el sitio web. Estaríamos hablando básicamente de capacidad de búsqueda y recuperación de la información.

¹ Este artículo forma parte de los resultados del proyecto de investigación “BIBLIOTECA DIGITAL OVIDIANA: Las ediciones ilustradas de Ovidio. Siglos XV al XIX (III): Las bibliotecas de Castilla y León” (HAR2011-25853), financiado por el ministerio de Economía y Competitividad.

- **Confiabilidad.** Podríamos definir este término como ausencia de errores. La confiabilidad de un sitio web está relacionada tanto con la correcta ejecución de la funcionalidad proporcionada por el sitio como con el correcto procesamiento de los vínculos.
- **Eficiencia.** Hablamos del uso adecuado de los recursos. Puede verse en el tiempo de respuesta y la velocidad de generación de las páginas y gráficos así como en el buen uso del espacio de almacenamiento en el servidor.
- **Facilidad para recibir mantenimiento.** Está asociada con la fácil corrección, adaptación y extensibilidad del sitio web.

Todas estas características, junto con otras mencionadas por Offutt (Offutt, 2006) como es la escalabilidad, han sido tenidas en cuenta a la hora de enunciar los requisitos a satisfacer por la biblioteca digital ovidiana.

La facilidad para el uso está asociada con características como sencilla comprensión global del sitio web y la estética de la interfaz pero también, en el contexto de la biblioteca digital, con la agilidad para acceder a la información que se desea consultar. Estos elementos se han cuidado especialmente al diseñar la página, buscando un justo equilibrio entre estética y funcionalidad que compatibilice la usabilidad con la potencia de las funciones de búsqueda y consulta de la información. Para este fin se han diseñado búsquedas básicas por los principales criterios, siglos y comunidades autónomas, así como una búsqueda avanzada que permite localizar una subcadena en cualquiera de los campos. También se ha implementado un listado de ediciones por siglo a partir del que se obtienen todos los ejemplares de cada edición. La interfaz, basada en la ficha de los ejemplares, resulta muy homogénea e intuitiva. Esta sencillez facilita la consulta del contenido ya que, suministrando funciones para la navegación ágil entre páginas y funciones, no distrae del objetivo último de mostrar de una forma comprensible y bella la información de los ejemplares y sus ilustraciones.

La confiabilidad se ha cuidado mediante la realización de pruebas y también de una forma preventiva. Para ello se ha llevado a cabo un diseño de calidad y se han empleado los estándares propuestos por la World Wide Web Consortium (W3C) en la implementación; HTML5, estilos CSS adaptativos, y tecnologías para las aplicaciones web. Estas características del software lo hacen también más fácil de mantener, una necesidad de este tipo de sistemas cuyo desarrollo se extiende de forma continuada en el tiempo.

En cuanto a la eficiencia, se ha estudiado especialmente la velocidad de carga de las imágenes. Para ello se han creado archivos con imágenes reducidas (thumbnails) para las consultas en listas y se ha reducido el peso de los archivos de las ilustraciones adaptándolas a la precisión requerida en función del marco en que se van a visualizar.

La escalabilidad es uno de los requisitos básicos de la aplicación al tratarse de un proyecto en evolución. Inicialmente se digitalizaron los ejemplares de en las bibliotecas de Galicia y Cataluña pasando posteriormente a los de Castilla León. En próximas fases se digitalizarán los de las bibliotecas de las comunidades autónomas restantes. Por esto se ha puesto especial cuidado en el diseño de la base de datos en su normalización lo que permitirá, gracias a la inexistencia de información redundante, un mejor aprovechamiento del espacio de almacenamiento y una gestión más ágil y sencilla de la información. Esta ha sido también una de las razones por las que se ha reducido el peso de las imágenes. Sus archivos se han nombrado con una codificación articulada compuesta por el código del

ejemplar y el número (1, 2...) o la posición (p portada, ap anteportada...) de la imagen. Además se han organizado los archivos en carpetas por ejemplar. De esta manera resulta más sencilla la búsqueda de los archivos y su manejo y se facilita el crecimiento de los fondos digitalizados.

Pero ¿por qué no ser más ambiciosos? La biblioteca digital puede convertirse en una herramienta realmente potente de apoyo a la investigación que proporcione servicios de altísimo valor añadido que, haciendo uso de técnicas de inteligencia artificial, guíen al investigador en su búsqueda y le permitan encontrar incluso aquello que el desconoce que existe. ¿Por qué limitar el acceso a sus recursos a los usuarios del sitio web? ¿Por qué no abrir su contenido para hacerlo accesible desde servicios generales de búsqueda o desde aplicaciones que proporcionan servicios basadas en información recogida en muchas fuentes diferentes? Nuestros siguientes pasos deberán estar encaminados a dar respuesta a estas inquietudes.

Un diseño guiado por el contenido

La biblioteca ovidiana es un caso atípico de biblioteca digital. Por regla general una biblioteca digital es una colección de volúmenes digitalizados junto con una breve información descriptiva de los mismos, parte de ella estructurada, como son los campos título, autor, edición, fecha... Nuestro caso es diferente ya que por un lado no se han digitalizado completos los ejemplares sino sólo sus ilustraciones. Se ha realizado un estudio pormenorizado de cada uno de los ejemplares y de su contenido iconográfico. Este estudio da como resultado gran cantidad de información estructurada junto con las imágenes por lo que nos encontramos frente a un sistema con una compleja estructura de información. En esta información radica la clave del éxito. Por ello el trabajo de diseño del sitio web ha girado en torno al contenido y su arquitectura.

La base para el modelado de los datos fue la ficha bibliográfica elaborada por las expertas. A partir de ella, y en función de los criterios de las búsquedas, se elaboró el modelo conceptual de datos, modelo Entidad-Relación², que se representa en la figura 1 (Modelo conceptual de datos). Este tipo de modelos resultan muy útiles para el diseño, dado su naturalidad para el modelado, al representar el mundo en términos de

- entidades, objetos del mundo real sobre los que nuestro sistema guarda información,
- atributos, o propiedades de estas entidades, y
- relaciones, vínculos entre las entidades inherentes a la semántica de los datos en el mundo real.

Para la creación de la base de datos el modelo conceptual ha de traducirse en un modelo de representación³. En la implementación del sitio web <ovidius-pictus.com> se ha empleado un modelo lógico relacional de datos cuyo esquema puede verse en la figura 2 (Modelo lógico relacional). El esquema muestra una

² El modelo Entidad-Relación propuesto por Chen (1976) es la herramienta más extendida para el modelado conceptual de datos dada su naturalidad para diseñar modelos.

³ Los modelos de datos de representación muestran un mayor número de detalles que los modelos conceptuales sin llegar con ello a mostrar detalles físicos sobre cómo los datos son almacenados ni sobre los caminos de acceso a los mismos.

estructura con grandes semejanzas a los modelos en copo de nieve empleados para modelar almacenes de datos corporativos⁴ con los criterios de búsqueda a modo de dimensiones y las ediciones y ejemplares, en el corazón del modelo, en paralelismo con las tablas de hechos. Esta semejanza calculada facilita las tareas de recuperación de la información que, al igual que en los modelos para la inteligencia de negocio, es el objetivo principal de nuestro sistema.

El diseño de la arquitectura del sitio sigue una estructura básicamente jerárquica con lo que se combinan una navegación muy intuitiva y una gran agilidad.

Se ha incluido, junto con el menú de opciones y los botones de navegación para poder avanzar y retroceder en las imágenes..., una barra de estado que informa en todo momento del significado de la información presentada. El ejemplar, núcleo de nuestro modelo de datos, es también el elemento central de la navegación y la barra de estado muestra el ejemplar activo para poder volver fácilmente a él.

EL NUEVO HORIZONTE: LA BIBLIOTECA DIGITAL OVIDIANA ABIERTA

El proceso de digitalización de contenidos bibliográficos cumple con un doble cometido. El primero la perdurabilidad, haciendo que estos contenidos no desaparezcan por el deterioro del soporte papel en el que están impresos. El segundo objetivo es hacerlos accesibles a un gran número de usuarios con el objeto de sacar el máximo rendimiento de los mismos. Esta disponibilidad se incrementa al publicarlos en la web. La Biblioteca Digital Ovidiana cumple con estos dos objetivos proporcionando una herramienta funcional y amigable mediante la que es posible acceder a la información sobre ediciones ilustradas de Ovidio en las bibliotecas españolas, a la información específica de cada uno de los ejemplares y al contenido iconográfico de los mismos.

Pero, al igual que en el ámbito profesional se llevan a cabo procesos de inteligencia de negocio en los que se emplea todo el potencial de los datos disponibles en la organización para sustentar la toma de decisiones de gestión, estos datos pueden ser empleados de forma eficiente para, mediante la investigación, extraer de ellos todo su potencial. Para ello se debe, por un lado abrir la biblioteca para incrementar la visibilidad de los datos y dotarle de contenido semántico para hacer más eficiente y automático su tratamiento al incluir junto con los objetos digitales información que posibilita a las máquinas entender su contenido.

De acuerdo con estos objetivos se está llevando a cabo el proceso de apertura, para hacer accesible su contenido a través de buscadores y otras interfaces, e integración de la biblioteca digital ovidiana en Hispana, con el fin de que sus datos puedan ser recolectados tanto por ella como por Europeana. De esta forma se incrementará la visibilidad de los datos. Además, al hacer públicos los metadatos y con el enriquecimiento semántico de los mismos se conseguirá no sólo perfeccionar los resultados obtenidos por los buscadores sino la posibilidad de implementar servicios de alto valor añadido capaces de sacar todo el partido a la riqueza de la información que guarda la biblioteca.

⁴ Más conocidos por su nombre en inglés, *Data Warehouse*. Puede verse Chaudhuri (1997) para obtener más información.

El recolector Hispana dentro del ámbito del proyecto Europeana

Europeana es un proyecto europeo que comenzó en 2005 y cuyo objeto es centralizar en una gran biblioteca digital el acceso a contenidos culturales procedentes de diferentes instituciones de países de la unión europea lo que constituirá "el punto común de acceso a los fondos de las bibliotecas, archivos y museos de toda Europa" (Comisión Europea, 2011).

En cuanto a la organización de Europeana, esta no almacena contenidos sino que se encarga de recopilar conjuntos de metadatos que describen los recursos. La responsabilidad de los documentos se organiza en tres niveles:

- Las bibliotecas digitales asociadas son las responsables del mantenimiento de los documentos y sus metadatos.
- Un agregador, a nivel estatal generalmente, agrega los datos de las bibliotecas.
- Europeana recopila los datos de los agregadores nacionales y los hace disponibles a los usuarios. Cuando un usuario quiere recuperar un documento le remite a la biblioteca de origen.

Europeana centra su actuación en cuatro líneas estratégicas (Europeana, 2012):

1. Agregar contenidos para construir una fuente abierta de patrimonio europeo confiable.
2. Facilitar la transferencia de conocimiento, la innovación y la promoción de la herencia cultural.
3. Distribuir su patrimonio libremente a los usuarios estén donde estén, siempre que lo deseen.
4. Atraer a los usuarios en las nuevas formas de participación mejorando la experiencia y promoviendo la comunicación entre los conservadores/gestores de los objetos digitales y usuarios mediante el uso de herramientas de la web 2.0.

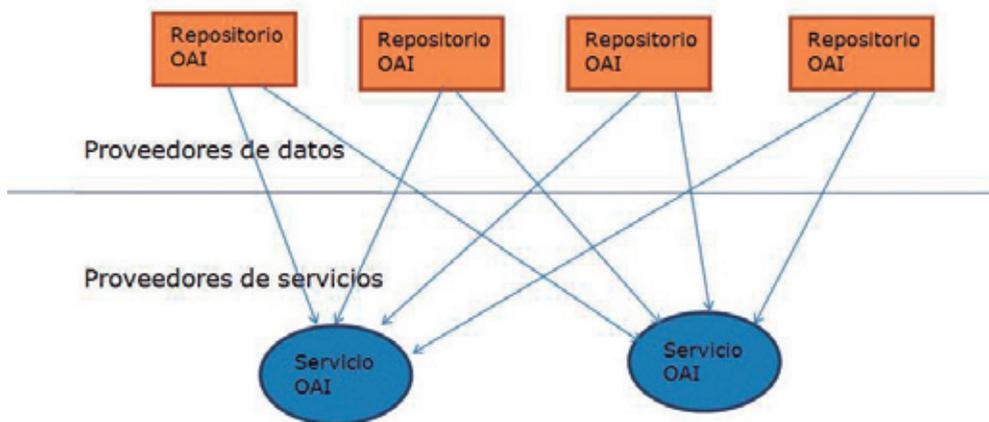


Fig. 3. Arquitectura básica del protocolo OAI-PMH.

Hispana reúne las colecciones digitales de archivos constituyendo un agregador de contenidos de las bases de datos de colecciones digitales empleando el protocolo OAI-PMH (Barrueco, 2003).

Este protocolo de interoperatividad permite la recopilación de metadatos descriptivos basados en Dublin Core⁵ en formato XML⁶. Según este protocolo existen dos tipos de agentes:

- Proveedores de datos (data providers). Gestionan los objetos y metadatos asociados a través de repositorios digitales y los ponen a disposición de otros sistemas vía OAI-PMH.
- Agregadores (harvester): Recopilan los metadatos provenientes de diferentes proveedores de datos y los almacenan en una base de datos que posteriormente puede ser consultada por aplicaciones web o usuarios finales. Todo ello es gestionado por lo que se conoce por un proveedor de servicio.

El intercambio de metadatos posibilita el desarrollo de buscadores que indexen los contenidos de recursos digitales permitiendo a los usuarios acceder a información de diferentes repositorios en una sola consulta.

Hispana realiza la función de harvester recopilando datos de distintas bibliotecas digitales que hacen la función de data providers. Por su parte Europeana, como ya se comentó, es un harvester especial que recolecta datos de los distintos recolectores nacionales realizando así una labor de centralización.

Enriquecimiento semántico de los metadatos según las directrices de la web 2.0

La web, tal y como se concibió originalmente, consiste en un conjunto de páginas con texto conectadas entre sí mediante hipervínculos. La información contenida en ella resulta comprensible para las personas que la consultan pero no así para las máquinas que sólo ven en ella una secuencia de palabras sin estructura. Es por ello por lo que cuando queremos localizar una información mediante un buscador este nos devuelve simplemente páginas en las que aparecen los términos que buscamos y no aquellas en las que se encuentra el significado de lo que queremos consultar. Si tuviéramos, por ejemplo, la figura 4 (Consulta de un ejemplar como texto plano) la página aparecería entre los resultados de una búsqueda de "Metamorfosis Tarragona" y la ilustración de la portada en una búsqueda de "Imágenes Tarragona".

En Codina y Rovira (2006) se dan dos definiciones complementarias de web semántica⁷

"Definición 1. La visión de la Inteligencia Artificial: La web semántica es un conjunto de iniciativas destinadas a promover una futura web cuyas páginas estén organizadas, estructuradas y codificadas de tal manera que los ordenadores sean capaces de efectuar inferencias y razonar a partir de sus contenidos.

⁵ Dublin Core: Modelo de metadatos orientado a la descripción de recursos digitales.

⁶ XML: eXtensible Markup Language. Es un metalenguaje para la definición de vocabularios orientados al intercambio de datos estructurados.

⁷ En García (2005) puede verse una visión general de esta tecnología.



Fig. 4. Consulta de un ejemplar.

Definición 2. La visión del procesamiento robusto: La web semántica es un conjunto de iniciativas destinadas a convertir la World Wide Web en una gran base de datos capaz de soportar un procesamiento sistemático y consistente de la información”.

La Web Semántica (Fernández, 2008) es por tanto la nueva web comprensible no sólo por los humanos sino por sistemas inteligentes. En ella se incluyen anotaciones que describen formalmente la información, el conocimiento, que contiene. Gracias a esto será posible desarrollar sistemas que exploten este conocimiento para inferir nueva información útil para los usuarios. Las anotaciones antes citadas son los metadatos.

Según la norma ISO/UNE 23081 (ISO, 2012) “los metadatos se definen como datos que describen el contexto, contenido y estructura de los documentos, así como su gestión en el tiempo”.

Los metadatos de una biblioteca digital son la información acerca de la estructura de los datos almacenados en la misma. Según definen Hernando-De-Larramendi (2009) “aquella información que permite la descripción pormenorizada de copias digitales y de sus originales, los requisitos técnicos relacionados con la captura de las imágenes o generación de textos y el almacenamiento de los mismos y sus características, y la descripción de todos y cada uno de los procesos que dan lugar a los objetos digitales definitivos”.

Existen distintos tipos de metadatos para la gestión archivística (Rosado Pacheco y Enrique Muñido):

1. Metadatos administrativos. Para la gestión de recursos de información.
2. Metadatos descriptivos. Sirven para representar o identificar los objetos de información digital.
3. Metadatos de conservación. Destinados a gestionar la preservación de los elementos de información.
4. Metadatos técnicos. Para la gestión automática de los objetos.
5. Metadatos de uso. Relativos al nivel de utilización de los servicios de información.

Para poder ser recolectados, los metadatos deberán codificarse empleando un lenguaje, modelo, de metadatos, común. Para el proyecto Europeana se diseñó un modelo de metadatos unificado denominado ESE en el que se definían, mediante un conjunto de etiquetas, atributos como la descripción, el autor, el editor, el formato o el tipo de los objetos digitales. Este modelo, fácil de aplicar por su simplicidad, se ha comprobado que es insuficiente dada su falta de flexibilidad y su incapacidad para describir aspectos semánticos propios del contenido de los objetos de información. Esto, junto con su estructura plana que imposibilitaba la inclusión de enlaces externos y la extensibilidad a modelos más especializados, ha motivado su remplazo por el nuevo modelo de Europeana denominado European Data Model (EDM). EDM amplía las posibilidades de búsqueda y posibilita el enriquecimiento semántico de los datos soportando objetos más complejos. Según Doerr (2010) "Ofrece una mayor expresividad y flexibilidad y permite representación más rica y verdaderamente semántica de los millones de objetos de todo tipo de las comunidades de patrimonio cultural en Europeana".

Por otra parte, la Web Semántica no sólo necesitan tener acceso a los datos, también debe estar disponibles las relaciones entre ellos. Hay que crear una web de datos en lugar de una mera colección de conjuntos de datos (W3C). Esta colección de conjuntos de datos relacionados entre sí en la web también se conoce como Linked Data. Los Uniform Resource Identifier (URI), cuyo subconjunto más conocido son los URL, son un mecanismo para identificar de forma inequívoca cualquier recurso en esta red. En la Web Semántica juegan también el rol de identificadores de objetos del mundo real, físicos y no de información (Peis, 2003).

EDM, para su integración en la iniciativa Linked Data, se sustenta en la combinación de varias tecnologías o estándares para compartir datos estructurados en la web (Doerr, 2010).

- El estándar RDF (Resource Description Framework), permite representar información estructurada sobre cualquier recurso mediante triples con la forma (sujeto, predicado, objeto). Los objetos se identifican mediante URIs⁸ sobre las que se declaran propiedades y se establecen relaciones. Estas propiedades junto a los tipos de sujetos y objetos se definen en las ontologías⁹. RDF es el lenguaje de representación de la Web Semántica y el metamodelo adoptado por EDM.
- Las ontologías se definen por medio de RDF Schema (RDFS) y del estándar Web Ontology Language (OWL).
- EDM reutiliza algunas ontologías de referencia como Simple Knowledge Organization System (SKOS) que proporciona elementos para representar un sistema estructurado de conocimiento como pueden ser tesauros, clasificadores o listas de encabezados de materias.
- Dublin Core (DC) proporciona términos para describir las características esenciales de los objetos culturales (creadores, relaciones con otros recursos, índices por materias, etc.). Es un modelo sencillo que facilita además la reutilización de los datos en EDM ya que DC está muy extendido y lo usan muchas aplicaciones.

⁸ Un *Uniform Resource Identifier* o *URI* (en español "identificador uniforme de recursos") es una cadena de caracteres corta que identifica inequívocamente un recurso en la web.

⁹ Empleamos el término ontología como sinónimo de conceptualización.

- Friend-Of-A-Friend (FOAF): es una ontología pensada para describir personas y sus perfiles y se emplea fundamentalmente en redes sociales.
- OAI Object & Exchange (OAI-ORE): es el marco estructural para definir conjuntos de recursos en la red así como recursos complejos formados mediante agregación de otros más sencillos. Mediante OAI-ORE es posible asociar un objeto con diferentes instancias del mismo representadas mediante un conjunto de recursos presentes en distintas ubicaciones.

La información de la consulta antes comentada se podría representar con la estructura de metadatos mostrada en la figura 5 (Grafo RDF ejemplo).

El archivo XML del ejemplo anterior sería:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:j.0="#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:edm="http://www.europeana.eu/schemas/edm/"
  xmlns:gss="http://www.w3.org/2001/11/IsaViz/graphstylesheets#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:ore="http://www.openarchives.org/ore/terms/"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:op="http://www.ovidiuspictus.es/terminos#"
  xmlns:xhtml="http://www.w3.org/1999/xhtml/vocab#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="file:/C:/Users/Paloma/Documents/documentos/investigación/ovidio/hispana/ejemxml">
  <op:comunidad rdf:about="op:comunidad_1">
    <rdf:object>http://viaf.org/viaf/170986710</rdf:object>
    <dc:title>Cataluña</dc:title>
  </op:comunidad>
  <rdf:Description rdf:about="op:listaobras_99">
    <rdf:li>
      <op:obra rdf:about="op:obra_1">
        <rdf:object>http://viaf.org/viaf/180872106</rdf:object>
        <rdf:value>Metamorfosis</rdf:value>
      </op:obra>
    </rdf:li>
    <foaf:isPrimaryTopicOf rdf:resource="file:/C:/Users/Paloma/Documents/documentos/investigación/ovidio/hispana/ejemxml#obra"/>
  </rdf:Description>
  <rdf:Description rdf:about="op:portada_103">
    <edm:object>M.BPT.Ven.1509.p.jpg</edm:object>
    <edm:isShownAt>http://www.ovidiuspictus.es/images/images/M.BPP.Par.1802.t4/M.BPP.Par.1802.t4.p.jpg</edm:isShownAt>
    <edm:type>IMAGE</edm:type>
  </rdf:Description>
  <op:biblioteca rdf:about="op:biblioteca_20">
    <op:autonomia rdf:resource="op:comunidad_1"/>
    <dc:title>Tarragona. Biblioteca Pública de Tarragona</dc:title>
  </op:biblioteca>
  <op:edicion rdf:about="op:edicion_104">
    <dc:title>Accipe Studiose Lector<br />P. Ouidii Metamorphosin cum luculentissimis<br />Raphaelis Regii enarrationibus: quibus<br />plurima as-
```


cripta sunt: que in exempla
ribus antea impressis non inveni
untur. Que sint rogas: Inter
legendum faciletibi
ocurrent.
Cum Gratia et Privilegio.

Colofón:
Ad lectorem
Siquid forte litterar[um] immutatione. transpositione. inuersione
appositione. omissione aliaue depravatio[n]e offenderis stu
diose lector id correctionis difficultati ascribas ro
gat Georgius de Rusconibus Mediolane[n]sis
cuius industria Raphael Regius i[n] hoc
opere describendo usus est. Ve
netiis Principe felicis.
Leonardo Lau
reda[n]o die. ii.
maii. M.D.
IX.</dc:title>

```
<op:codigoedicion>Metamorfosis.Regius.Rusconi.Venecia.1509</op:codigoedicion>
<op:obra rdf:resource="op:listaobras_99"/>
<dc:date>1509</dc:date>
</op:edicion>
<op:ejemplar rdf:about="op:edicion_103">
  <op:edicionejemplar rdf:resource="op:edicion_104"/>
  <op:imagenportada rdf:resource="op:portada_103"/>
  <op:bibliotecaejemplar rdf:resource="op:biblioteca_20"/>
  <foaf:useGuidelines>M.BPT.Ven.1509</foaf:useGuidelines>
</op:ejemplar>
</rdf:RDF>
```



Fig. 6. Estructura evolutiva del proceso.

Fases del proyecto

Con el fin de implantar un modelo de datos acorde con las tecnologías de Web Semántica y Linked Data para la integración de la biblioteca digital ovidiana a Europea se ha diseñado una planificación que se puede ver en figura 6 (Estructura cíclica del proceso).

Con el fin de garantizar el éxito del proyecto se ha optado por un modelo de proceso en espiral en el que el modelo irá refinándose en cada una de las iteraciones. Las iteraciones que se planean realizar son las siguientes:

En este momento nos encontramos en la primera iteración, describiendo los metadatos descriptivos. Para ello se ha definido el metamodelo a partir del modelo conceptual de datos. Se ha representado una instancia de ejemplar según el anterior modelo de metadatos. Es la correspondiente al M.BPP.Par.1802.t4, una edición de las Metamorfosis de París del siglo XIX que tiene cuatro tomos y se encuentra en la Biblioteca Pública de Pontevedra. Esta instancia deberá ser validada. En las próximas iteraciones del proceso de desarrollo se irán incluyendo los metadatos técnicos, administrativos, de gestión y de uso.

BIBLIOGRAFÍA

- Barrueco, J.M., "OAI-PMH: protocolo para la transmisión de contenidos en Internet", *El profesional de la información*, 12, 2 (2003), pp. 99-106.
- Chaudhuri, S., "An overview of data warehousing and OLAP technology", *ACM Sigmod record*, 26, 1 (1997), pp. 65-74.
- Chen, P., "The Entity-Relationship Model: Toward a Unified View of Data", *ACM Transactions on Database Systems (TODS)*, 1, 1 (1976), pp. 9-36.
- Codina, L. y Rovira, C., *La Web Semántica*, 2006.
- Comisión Europea, *European's Information Society. Thematic portal*, 27 de octubre de 2011, Iniciativa sobre Bibliotecas Digitales:: Europea: <http://ec.europa.eu/information_society/activities/digital_libraries/europeana/index_es.htm> [recuperado el 25 de junio de 2013].
- Doerr, M.G., "El Modelo de Datos de Europea (EDM)", *Meeting: 149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management*, Gotemburgo, Suecia, 2010: WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY.
- Europeana, *Strategic Plan 2011-2015*, 2012, <http://www.pro.europeana.eu/c/document_library/get_file?uuid=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602> [Recuperado el 28 de junio de 2013].
- Fernández, R.C., *Representación del Conocimiento. Web Semántica*, Madrid, Universidad Carlos III de Madrid, 2008.
- García, N.F., "La Web Semántica: fundamentos y breve estado del arte", *Novática: Revista de la Asociación de Técnicos de Informática*, 178 (2005), pp. 6-11.
- Hernando-De-Larramendi, L.D.-M.-P.-C., "Datos y metadatos: la normalización dinámica de los elementos y de los procesos constituyentes de una Biblioteca Virtual", *FESA-BID'09*, 2009, pp. 109-116.
- ISO, *UNE-ISO/TR 23081-3 Metadatos para la gestión de documentos*, 2012.
- Offutt, J., "Quality Attributes of Web Applications", *IEEE Software*, 19, 2 (2002), pp. 25-32.
- Olsina, L.E., "Specifying Quality Characteristics and attributes for Web Sites", *Proc. 1st. ICSE Workshop on Web Engineering*, Los Angeles, ACM, 1999.
- Peis, E.H.-V., *Ontologías, metadatos y agentes: recuperación "semántica" de la información*, 2003.

Pressman, R.S., *Ingeniería del Software. Un enfoque práctico*, 7ª ed., México D.F., McGraw Hill, 2010.

Rosado Pacheco, I. y Muñido, Enrique J., "El viaje de los documentos: Del anaquel a la nube. Metadatos y principales repositorios de objetos digitales", *Códice. Gestión de la información*, s.f.

W3C, *Semantic Web*, Linked Data, s.f., <<http://www.w3.org/standards/semanticweb/data>> [Recuperado el 2 de julio de 2013].



RESUMEN

La *Biblioteca Digital Ovidiana* es un caso atípico de biblioteca digital que tiene un sistema con una compleja estructura de información y que cumple con el doble cometido de la perdurabilidad y disponibilidad de la información sobre las ediciones ilustradas de Ovidio en las bibliotecas españolas. Pero con el objeto de extraer de estos datos todo su potencial, se ha dispuesto abrir la biblioteca para incrementar la visibilidad de los datos y dotarla de contenido semántico, para hacer más eficiente y automático su tratamiento al incluir junto con los objetos digitales la información que posibilita a las máquinas entender su contenido. De acuerdo con estos objetivos, se está llevando a cabo el proceso de apertura, para hacer accesible su contenido a través de buscadores y otras interfaces, y de manera concreta para integración de la biblioteca en *Hispana*, con el fin de que sus datos puedan ser recolectados tanto por ella como por *Europeana*. De esta forma se incrementará la visibilidad de los datos y, al hacer públicos los metadatos y con el enriquecimiento semántico de los mismos, se conseguirá no sólo perfeccionar los resultados obtenidos por los buscadores, sino la posibilidad de implementar servicios de alto valor añadido capaces de sacar todo el partido a la riqueza de la información que guarda la biblioteca.

Palabras clave: Biblioteca Digital, Europeana, Web Semántica, Calidad Web.

ABSTRACT

The *Biblioteca Digital Ovidiana*, which meets the double mission of providing durability and availability to the information about the Ovid illustrated editions kept in Spanish libraries.

However, it has been arranged to become an open library, in order to extract the full data potential and be enable to also increase the data visibility. Additionally, the semantic content, which includes digital objects information that enables machines to understand its content, allows automatic data treatment. According to these goals, the opening process is being carried out to make the content accessible through search engines and other interfaces, including the integration in *Hispana*, and so that the data could be also harvested by *Europeana*. This will increase the visibility of the data and, by making public the metadata and by their semantic enrichment, will be improved not only the results obtained by the search engines, but also the possibility of implementing value-added services, capable of take full advantage of the rich information stored in the library.

Keywords: Digital Library, Europeana, Semantic Web, Web Quality.

Tecnoestética de la blogoficción: imitación y suplantación del bloguero en la narrativa del avatar

DANIEL ESCANDELL MONTIEL
Universidad de Salamanca
danielescandell@usal.es

El espacio social digital ha albergado toda una línea creativa literaria en forma de escrituras de la red asociadas al blog en todas sus vertientes, esto es, desde la bitácora en su sentido más amplio hasta los sistemas de microblog y nanoblog¹ que se asocian al concepto de red social que representan sistemas como Twitter o Facebook.

La blogoficción es un entorno abierto bajo el que se han amparado diferentes estilos literarios que incluyen composiciones poéticas, narrativas, etc. No en vano, se trata de una plataforma de comunicación flexible y abierta que puede ser empleada como herramienta de publicación y distribución textual en red. Sin embargo, resulta evidente que algunos modelos textuales han florecido con especial fuerza y han formado un estilo propio y definido que se siente en un estado natural en la plataforma blog. Nos referimos a los textos narrativos homodiegéticos, aquellos en los que el autor encarna el papel de un bloguero para construir su historia y orienta su técnica literaria a la asimilación de los usos y costumbres de la red para asumir ese rol en el blog, en la red social, y pretender ser un internauta que cuenta a la comunidad digital su vida. Se trata de la emulación del bloguero, de un proceso de pretensión de lo real, mediante la imitación de la estética de esos espacios. Si el blog nace como diario público, la blogoficción adopta ese formato y su uso mayoritario para componer la obra fictiva. Estas creaciones son, asimismo, alargadas en el tiempo (no pocas se han extendido durante meses e incluso años) en un cronotopo de la realidad: el tiempo de la blogoficción es el del mundo real y, por tanto, es el que comparten

¹ Frente al nanoblog, el microblog permite textos de mayor tamaño que pueden estar limitados por el sistema, pero en todo caso de mayor extensión. Un formato de publicación no restringido por el sistema, aunque sí por el uso generalizado de los usuarios, es Tumblr. Típicamente están orientados a compartir contenido multimedia o textual, aunque de mayor extensión, pero en ningún caso se espera de ellos mensajes extensos. Como en el nanoblogueo, el microblogueo se integra también en sistemas de redes sociales, ya que estas se orientan principalmente a la publicación de textos de extensión pequeña o comedida y contenidos multimedia de diverso tipo.

autor, lectores y personajes. Se trata de una categoría compartida e inalterable (aunque algunas plataformas de bitácoras pueden permitir la falsificación de las fechas de publicación) que conforma un presente absoluto en la obra blogofictiva en el que se sitúan los diferentes capítulos (entradas de blog, mensajes de red social, etc.) ordenados, cómo no, siguiendo la estructura cronológica inherente a nuestra percepción lineal del tiempo.

En la blogoficción el autor emplea un canalizador, un ente demiúrgico (Baudrillard, 1978: 7-80) que es el avatar, esto es, el protagonista narrador de la obra que es, a su vez, el bloguero. A través del avatar se construye una doble realidad impostada: por un lado, la de la ficción literaria que el autor está componiendo (y que no tiene que ser asumida como tal por los lectores, pues uno de los objetivos que se persiguen es el del simulacro y, por tanto, el *hoax*² ante los receptores); por otro, la de la ficción vital fuera de la obra que se deriva del avatar. Y es que este avatar debe desbordarse de la obra para cubrir el paratexto, lo que en este sentido implica relacionarse con los lectores en los comentarios de los blogs, abrazar otros espacios sociales de la red para ocuparlos y constituirse como *persona* también en ellos y, en definitiva, vivir no sólo en la obra –el blog– sino también fuera de él para que la simulación sea completa. Este bloguero que encarna el autor literario mediante el avatar no es un personaje de la bitácora, sino un personaje de la red: habita en ella (y, claro, en el mundo real) y, al mismo tiempo, nos habla a nosotros desde su blog. Sólo así la simulación es plena y el avatar puede cumplir con su objetivo de impostación.

El presente compartido y la realidad compartida (mientras se mantiene el *hoax* de la blogoficción, la realidad fictiva debe ser correspondiente con la realidad cognoscible mediante nuestro conocimiento del mundo como habitantes del mismo) ayudan a conformar el rizoma de una memoria histórica de la web misma, sobre todo en cuanto a la concepción cronotópica de internet que se ha impuesto en los formatos de publicación mayoritarios (red social, blog, etc.), donde “la existencia de enlaces permanentes y archivos cronológicos constituye una de las contribuciones más importantes de los weblogs a la web, pues le han aportado memoria histórica” (Orihuela, 2006: 48). La memoria no sólo es compartida, sino que se puede reconstruir cronológicamente pese a que el dinamismo en la aportación de contenidos siempre ha existido en los diferentes formatos web. La memoria colectivizada, compartida públicamente en los entornos digitales, constituye un nuevo espacio *virtual* en el que se crea una memoria social complementaria y sumativa a la orgánica apoyada en los mecanismos de almacenamiento y distribución de la información que pueden emplear los usuarios gracias a los instrumentos y dispositivos de la digitalidad, especialmente cuando estos establecen la conexión permanente a la red mediante las tecnologías de comunicación actuales: el nodo es permanente y cada persona se convierte en foco de emisión y recepción de información, en testimonio permanente del tiempo presente y eso, en definitiva, es futura huella del pasado, esta vez *fijada* en bytes.

² En referencia a la distribución de información falsa a través de internet, con el objetivo de divulgar lo máximo posible el engaño, hasta tal punto que algunos se han convertido en memes, o leyendas urbanas. Aunque pueden esconder intereses como la recolección de datos (cuentas de correo electrónico), o iniciar una campaña de publicidad viral, no es extraño que algunas nazcan con la simple intención de confundir a los crédulos.

Ese es uno de los juegos fundamentales de la blogoficción como resultado de la constitución en la red de una pluralidad de *yos* mediante el ente digital poliédrico del individuo en internet. El avatar es un proceso que puede ser tanto disociativo como una proyección en la que la propiocepción del individuo se desplaza desde la persona hasta la entidad virtual como adopción de *otros yo* digitales, lo que resulta en una fragmentación en el espacio-red dado que:

Muchas manifestaciones de la multiplicidad en nuestra cultura, incluida la adaptación de personalidades *online*, están contribuyendo a un replanteamiento general de las tradicionales concepciones unitarias de la identidad en este contexto, las experiencias con la comunidad virtual nos ayudan a elaborar estas nuevas visiones del yo (Turkle, 1998: 51).

El *yo* es parcial y fragmentado, se divide y multiplica en la proliferación de espacios digitales no vinculados porque tras cada *nick*, tras cada imagen icónica, hay *otro yo* que no tiene que ser, necesariamente, coincidente con los demás ni con el *auténtico yo*, aun cuando todos son –claro– lados diferentes de la misma compleja y poliédrica personalidad de la psique humana. El avatar es un retrato que puede ser visto “hacia adentro, hacia fuera, el rostro como la verdadera máquina del espíritu” (Brea, 1991: 106) que es capaz de exteriorizar lo interno-subjetivo de un lado del poliedro del individuo, o incluso varios de ellos, cuando no los totalmente simulados. Esto es posible también por la concepción lúdica de la relación social virtualizada a través del distanciamiento entre el referente real (*yo*) y los tecnocuerpos que empleamos como *proxy*³ en estas relaciones acercándonos, así, a “the utopian dream of liberating play from the game, of a pure play beyond the game, merely opened the way for the extension of gamespace into every aspect of everyday life” (Wark 2007: 169). El tecnocuerpo que se construye el individuo (sea este una imagen icónica, una foto que le representa, o un personaje completo de un videojuego –como *World of Warcraft*– o un entorno social en mundo sintético –*Second Life*–) es una proyección de su creador o, al menos, de parte de su personalidad, fragmentada y seleccionada para ese contexto social concreto. No es sólo una imagen, sino toda una serie de descriptores textuales a los que se suma una imagen proyectada que se puede conseguir por el uso de una fotografía o un personaje creado, tanto en un sistema social como en un videojuego u otro sistema de comunicación digital, y mucho más, pues también resulta ser parte del retrato cómo se comporta, cómo se comunica... en definitiva, su personalidad alterada (o segmentada) por el filtro digital del avatar como espejo deformante. Se forma con el avatar una “teleidentidad que no requiere de presencia física del sujeto sino de la concreción de las herramientas informáticas” (Tortosa, 2008: 262). Y es que aunque el tecnocuerpo es descrito por Echeverría como un “cuerpo humano implementado por un conjunto de prótesis tecnológicas que le permiten acceder y ser activo en el tercer entorno” (2003: 18), este no se constituye por la suplantación de elementos biológicos, sino por la suma e integración de los elementos tecnológicos que dan acceso a esta nueva esfera y el avatar no es sino una prótesis más que nos permite transportarnos a los mundos de bytes.

³ En este sentido, entendemos que el tecnocuerpo no es sólo una traslación de la identidad del usuario, sino una puerta de acceso a los mundos sintéticos que habita mediante sus proyecciones avatáricas. No es, en consecuencia, un mecanismo limitado a la *suplantación* de la persona en lo digital (un *surrogate*, si queremos), sino su vía de entrada a ese universo de bits.

En definitiva, el individuo escoge qué le representa como máscara virtual en estas interacciones y el avatar se presenta como un ente virtual que se convierte en rostro de un fragmento de la personalidad de su usuario. Es fruto de una combinación de atrezzo y de vestuario para ayudar a la escenificación digital del papel interpretado por el usuario-autor. Por eso es habitual encontrar comportamientos diferenciados de una misma persona en entornos digitales no interconectados (la máscara permite asumir ese *otro yo*). Como fruto de una estética del simulacro, la impostura avatárica debe ser completa en el estilo literario pues se ocupa por el ente virtual todo el espacio de inscripción que debería ser reclamado por el autor: el escritor no puede anexionarse ese espacio de reconocimiento de la autoría para constituir su presencia en la obra porque rompería la simulación. En este sentido, el avatar lo abarca y consume todo al construirse como un yo digital fruto de la fragilidad del estatuto del ego que, sin embargo, se desarrolla como consagración de este mismo. Es un narcisismo egotista que hace, a su vez, que cada blog sea diferente del resto de millones de bitácoras similares: su personalidad, única, es lo que inyecta ese valor distintivo. En un maremágnum de millones de blogs nos encontramos con que –en la práctica– todos los campos temáticos, los estilos y las funciones están ocupados y es difícil producir un espacio propio. Por eso la visión subjetiva de los temas tratados, la exposición vital, la narración de la experiencia propia, se convierte en el valor diferenciador y esencial de todos y cada uno. A su vez, se da una fuerte corriente comunal y, con ella, llega la remezcla, recombinación y redifusión de los contenidos en la blogosfera y en la red social.

La comunidad se basa, así, en la remezcla interna de los blogs y en la externa entre el conjunto de bitácoras, siendo esta la blogosfera que, en última instancia, se orientará a la simulación –como “funcionamiento del simulacro en cuanto maquinaria, máquina dionisiaca” (Deleuze, 1968: 263)– en la blogoficción al ser la bitácora una de las tecnologías del mundo del simulacro que identifica Graeber, puesto que “the technologies that have advanced since the seventies are mainly either medical technologies or information technologies – largely, technologies of simulation” (2012). Se genera aquí una “*estética de la apariencia digital*, la tecnoestética, que se vuelca con entusiasmo a la simulación y el simulacro, nublando la transparencia de cualquier verdad” (Marchán, 2006: 30).

Esta tecnoestética de la blogoficción es heredera de los casos de suplantación de identidad con intención distinta a la artístico-literaria. El antecedente por antonomasia, pues alcanzó gran popularidad en su momento, es el de Kaycee Nicole Swenson, una adolescente estadounidense que creó el popular blog *Living Colours* (actualmente desaparecido) con la ayuda de un amigo y que fue incluso entrevistada por el prestigioso *New York Times* el 10 de agosto de 2000. El problema es que Kaycee no era real, sino un perfil empleado por una mujer de mediana edad, Debbie Swenson, para crear a una joven deportista y enferma de cáncer que se convirtió en todo un símbolo en la red estableciendo una amplia red de contactos que creían que era real. Cuando Debbie decidió dar muerte a su personaje no se tardó mucho en unir cabos y la propia comunidad que la había convertido en un símbolo logró desentrañar la realidad tras Kaycee, algo que el rotativo antes citado fue incapaz de lograr (la entrevista se hizo telefónicamente y, en teoría, hablaron con Debbie, que se hizo pasar por Kaycee), y que no reconoció hasta un año más tarde (Hafner, 2001). Fue la investigación de miembros de la red y el análisis atento de sus gustos y referencias culturales lo que les llevó a sospechar que algo no cuadraba en la historia de Kaycee, lo

que Debbie confesó definitivamente en mayo de 2001. Se hizo público entonces que Kaycee no había existido jamás y Debbie alegó que era un personaje basado en tres personas reales que había conocido y que habían muerto por diferentes tipos de cáncer.

Conocido el caso, cabe preguntarse si esta proyección avatárica –Kaycee– respondía a una voluntad creativa con intencionalidad fictiva o si era una pulsión derivada de una enfermedad. No podemos negar que recoge aspectos que serán considerados relevantes para la definición de la blogonovela y parte de la blogoficción, destacando la creación de un personaje avatárico y la impostura mantenida hasta el desenlace. Con todo, la historia que se desveló tras descubrirse el engaño y el hecho de que la propia autora eliminara el blog poco después, parecen apuntar, más bien, a intenciones alejadas de la creación y exploración literaria o ficcional. Otra cuestión, por supuesto, es que, fuera de esa intencionalidad, este caso represente una nueva senda para la realización de obras autofictivas o de desarrollo de una literatura sustentada en el aprovechamiento de los parámetros de publicación y creación de identidades digitales. En esa misma línea es importante señalar que la autora “ni durante ni después de terminada la historia ha reclamado su autoría sobre el producto ni ha mostrado interés alguno en resaltar el carácter artístico de su performance” (Cleger, 2010: 352-353), aunque en sentido contrario se ha defendido el recurso a la suspensión *involuntaria* de la credulidad como parte del proceso de recepción del texto (Carroll, 2001). Quizá no hubo una intención explícita de crear literatura en el *fraude* de Swenson, aunque esa voluntad de asumir una impostura es, definitivamente, compartida con la blogoficción. Esta, a su vez, basa su relación con los lectores no en la firma de un pacto de ficción, sino en el poder nutrirse de su credulidad para la impostura y la simulación.

Kaycee fue un avatar creíble (una chica adolescente normal y corriente, al fin y al cabo), pero han sido muchas las blogoficciones sustentadas en avatares improbables o imposibles. Existe, de hecho, una extensa nómina de blogs imposibles que son tratados como narraciones homodiegéticas en las que sus autores-gestores son avatares tan dispares como Jesucristo, Elvis Presley o John Fitzgerald Kennedy. Un completo listado lo aporta el estudio realizado por Davidson (2006), quien recopila una serie de bitácoras imposibles creados por blogueros que encarnaban el papel de diferentes personajes históricos o famosos, como Abraham Lincoln, el Marqués de Sade o Enrique VIII, por poner unos ejemplos. El catálogo está centrado en la blogosfera anglosajona y presenta la dirección de cada blog. A cada bitácora recogida en el libro se le atribuye un tema principal que sirve como descripción de la misma, y aporta una entrada publicada en ese blog, aunque sin dar ningún tipo de fecha.

Estos ejemplos son convenientes porque son personajes que constituyen un imposible cronológico, de manera que el lector que decida leer esas bitácoras sabe más allá de toda duda posible que son obras de ficción: no hay intención de engaño, sino de pacto de ficción directo entre autor y lectores. El autor literario emplea el avatar de un personaje histórico para ofrecernos un retrato de su vida, un diario intimista de ese personaje, o cualquier otro tipo de historia en la que son ellos mismos los que usan la plataforma blog para narrarla. No se trata, por tanto, de suplantaciones de personalidad, aunque estas también se han dado, tanto en blogs como en redes sociales y algunas han terminado en denuncias e incluso en los tribunales, en ocasiones porque no se ha comprendido el *animus iocandi* con el que se habían creado y, en otras, porque el fin paródico se con-

fundía con una agresividad verbal amparada en el –consagrado pero no tan real como a veces se cree– anonimato de la red y la máscara avatárica. Podemos afirmar que la parodia, u homenaje, o usurpación del espacio de producción de presencia en redes sociales es relativamente habitual como resultado de múltiples intenciones, entre las que el humor es sólo una de ellas. En algunos casos, la suplantación ha sido tan convincente que diversos medios se han hecho eco de tuits publicados supuestamente por ciertas celebridades en redes sociales como Twitter, sólo para descubrir más tarde que en realidad esas no eran sus cuentas de usuario o que, simplemente, no tienen ningún tipo de perfil en las mismas.

El que, posiblemente, sea el caso más significativo de suplantación de identidad en la blogosfera (y, al mismo tiempo, un caso de obra basada en el pacto de ficción y no en el engaño) lo firma el autor de origen argentino Hernán Casciari con *El diario de Letizia Ortiz* (2004). Se trata de una blogoficción narrativa homodiegética en la que el personaje gestor del blog (esto es, el avatar que encarna el autor a la hora de narrar los hechos) no es sino la actual esposa del heredero al trono de España, Felipe (al que se refiere en las entradas del blog como Filíp [sic]). Esta bitácora narra hechos cotidianos próximos al anuncio del compromiso de la pareja: las historias resultan ricas en anécdotas íntimas caracterizadas por una gran abundancia de hipocorísticos, visiones irreverentes del núcleo familiar, etc. Por supuesto, sólo un lector improbablemente crédulo concedería credibilidad a esta blogoficción, de manera que el público desembarca en esta blogonovela con el conocimiento previo de que es ficción, un ejercicio de impostura literaria que recurre a las técnicas y los elementos característicos de la bitácora para construir una narración prolongada en el tiempo que apuesta por la comedia paródica. Se da aquí, además, una paradoja avatárica: no tiene sentido simular porque ya se sabe que es falso y, sin embargo, sí hay impostura avatárica, de tal modo que el autor no reclama la autoría y no hace uso de los espacios de inscripción (que están conquistados plenamente por los rasgos del avatar). Se emplea la simulación avatárica incluso cuando esta no tiene sentido porque es el canalizador de la tecnoestética blogofictiva.

Bien diferente es el caso de *Más respeto, que soy tu madre* (2003-2004), donde Casciari crea a una ama de casa anónima. Se trata de una bloguera de mediana edad a la que su hijo mayor introduce en la red y ella, siguiendo su consejo, va contando su día a día, lo que resulta en desvelar a los lectores una familia de lo más singular llena de aventuras y desventuras cotidianas. Se trata de un avatar ante el cual el lector no tiene que asumir necesariamente –sin información previa– que es falso: se trata de una historia ambientada en el mundo real, con un personaje que puede ser auténtico y que no es improbable que esté escribiendo en internet sus vivencias personales. En este sentido, se trataría de uno más de los muchos blogs nacidos como pulsión catártica para dar salida a las angustias de la vida, tal y como se haría en el tradicional diario personal, sólo que ahora es un acto público lanzado a la inmensidad de la red. Es decir, se trata de una ficción que imita también en su intención primigenia a los blogs más estereotipados. Su protagonista, Mirta Bertotti, es una mujer mercedina que lleva las riendas de una familia difícil de aguantar y son las situaciones delirantes que cuenta lo que hace que el lector deba decidir si se cree lo que lee o si afronta una perspectiva crítica. Y, en tal caso, deberá determinar a su vez si esta visión crítica para con el blog debe canalizarse en seguir leyéndolo como un ejercicio fictivo por parte de un autor desconocido (que podrá o no desvelarse) o bien alejarse de él al comprender que no es una *digirrealidad*.

Tanto en el caso de Kaycee como en el Letizia o el de Mirta, el *hoax* de los autores (con independencia de si sus objetivos son literarios u otros), la composición textual responde a la imitación del retrato del bloguero estereotipado, es decir, aquel que muestra una tendencia egotista que le conduce a sociedades microgrupales donde se desarrollan y concentran actitudes de difusión pública de la esfera privada (la llamada *extimidad*). La blogoficción es posible cuando el *falso yo* avatárico se convierte en encarnación plena de un tecnocuerpo digital, es decir, cuando todos los referentes de proyección en internet del *yo* señalan no al autor real, sino al personaje que es a su vez gestor del blog, narrador y protagonista de la historia que cuenta en la red. Se trata de una concepción teatralizante del papel del autor: el escritor pone en marcha un acto performativo y figurativo con el que se pretende engañar a los lectores, haciendo de la propia narración un ejercicio de *hoax*. La estética de la obra se subyuga, por tanto, a la *imitatio* del modelo bloguista, es decir, a la transformación de lo doméstico en público siguiendo la ordenación telepoliciana (Echeverría, 1994: 161) que a su vez deriva en la invasión de lo privado por lo tecnológico-digital público. Se da, así, una superación del umbral que marca la frontera entre lo íntimo, lo privado y lo exhibido, desdibujándose en la sociedad-red los ámbitos de las esferas sociales que se habían considerado tradicionales hasta el momento en la sociedad occidental. Cabe por tanto preguntarse si los blogs extimistas muestran la vida de sus autores son diarios sin particular interés o si son obras de arte producidas por los nuevos artistas de la era digital de la misma manera que las blogoficciones que los imitan sí resultan ejercicios fictivos que ponen en práctica una determinada estética de la red.

Como hemos visto, en estas blogoficciones el autor explora los límites de la credulidad de sus lectores, la narración y el personaje deben sustentarse en historias verosímiles y, por tanto, que *pueden ser posibles*, aun cuando sus planteamientos son fronterizos con lo improbable, quizá porque el autor no puede resistirse a explorar los límites de las reacciones del público receptor de su obra. Retomando los ejemplos antes expuestos de las creaciones de Casciari, resulta evidente que en cualquier caso (por muy rocambolesca que sea la narración) resulta más creíble la historia de una mujer con una familia desestructurada en un vecindario obrero de una ciudad argentina que la historia de Letizia Ortiz sacando a la luz las intimidades de la Casa Real española. Pero en todos los casos, tanto si el avatar es concebible como una persona real exponiendo su vida como si este es inconcebible porque choca frontalmente contra toda lógica y conocimiento del mundo por parte del lector, las herramientas empleadas por los autores son las mismas y el objetivo es idéntico: *hacer creer* que lo que se cuenta es o *puede ser* auténtico.

BIBLIOGRAFÍA

- Baudrillard, Jean, *Cultura y simulacro*, trad. Antoni Vicens, Barcelona, Kairós, 2007 [1978].
- Brea, José Luis, *Nuevas estrategias alegóricas*, Madrid, Tecnos, 1991.
- Carroll, Jon, "The sad yet untrue story of Kaycee", *San Francisco Chronicles*, 2001, <<http://www.sfgate.com/entertainment/article/The-sad-yet-untrue-story-of-Kaycee-3314994.php>> [20/09/2013].
- Casciari, Hernán, *El diario de Letizia Ortiz*, 2004, <<http://letizia-ortiz.blogspot.com>> [20/09/2013].

- Casciari, Hernán, *Más respeto, que soy tu madre*, 2003-2004, <<http://mujergorda.bitacoras.com>> [20/09/2013].
- Cleger, Osvaldo, *Narrar en la era de las blogoficciones*, Nueva York, The Edwin Meller Press, 2010.
- Davidson, Paul, *The Lost Blogs. From Jesus to Jim Morrison*, Nueva York, Warner Books, 2006.
- Deleuze, Gilles, *La lógica del sentido*, trad. Miguel Morey, Barcelona, Paidós, 1994 [1968].
- Echeverría, Javier, "Cuerpo electrónico e identidad", en D. Hernández Sánchez (ed.), *Arte, cuerpo, tecnología*, Salamanca, Ediciones Universidad de Salamanca, 2003, pp. 13-29.
- Echeverría, Javier, *Telépolis*, Barcelona, Destino, 1994.
- Graber, Dave, "Of Flying Cars and the Declining Rate of Profit", *The Baffler*, 2012, <http://thebaffler.com/past/of_flying_cars> [20/09/2013].
- Hafner, Katie, "A Beautiful Life, an Early Death, a Fraud Exposed", *New York Times*, 2001, <<http://www.nytimes.com/2001/05/31/technology/a-beautiful-life-an-early-death-a-fraud-exposed.html?pagewanted=all&src=pm>> [20/09/2013].
- Marchán Fiz, Simón, "Entre el retorno de lo real y la inmersión de lo virtual", en S. Marchán (ed.), *Real/Virtual en la estética y la teoría de las artes*, Barcelona, Paidós, 2006, pp. 29-59.
- Orihuela, José Luis, *La revolución de los blogs*, Madrid, La esfera de los libros, 1998.
- Tortosa, Virgilio, "Sujetos mutantes: nuevas identidades en la cultura", en D. Romero López y A. Sanz Cabrerizo (eds.), *Literatura del texto al hipermedia*, Madrid, Anthropos, 2008, pp. 257-272.
- Turkle, Sherry, "Repensar la identidad de la comunidad virtual", *El Paseante*, 27-28 (1998), pp. 48-51.
- Wark, McKenzie, *Gamer Theory*, Cambridge, Harvard University Press, 2007.



RESUMEN

La tecnoestética blogofictiva se define por el uso del avatar como mecánica narrativa que abre la puerta a todos los recursos de impostura y simulacro que las plataformas de publicación de tipo blog y social han abierto a los autores literarios. La adopción de personalidades *online* como resultado de la sociedad-red tiene su traslación en la creación literaria, que explota estos recursos en su propio beneficio para nutrirse de las plataformas digitales y para generar textos que exploran la frontera del pacto de ficción y el engaño al lector a través de la imitación de la figura del bloguero. Desde el caso de Kaycee Nicole Swenson, la inexistente joven que engañó al prestigioso *New York Times*, hasta las blogoficciones de Hernán Casciari (con personajes como Mirta Bertotti e incluso Letizia Ortiz), la blogosfera se ha convertido en un baile de máscaras en el que los autores juegan a la suplantación y conquistan, con sus personajes, los espacios personales de la red para explorar sus posibilidades literarias. Analizamos los casos señalados para dilucidar cómo ha influido la percepción y proyección del *yo* en la red a la hora de dar pie a una literatura nativa de blogs que, por tanto, es concebida para aprovechar estos espacios del mundo digital.

Palabras clave: Tecnoestética, blogoficción, avatar, *hoax*, red social.

ABSTRACT

The techno-aesthetics of blog-fiction may be described as the employment of an avatar as a narrative modality permitting access to the myriad possibilities for imposture and simulacrum that have become available to literary authors with the advent of blogging and

social networking platforms. The adoption of online personalities as a facet of the net-society has its counterpart in a sphere of literary creation that exploits the aforementioned resources in order to generate texts that test and even undermine the frontiers between fiction and hoax, utilizing the blogger's fabricated personalities. Since the case of Kaycee Nicole Swenson, the nonexistent teenage girl that duped the prestigious New York Times, to the blog-fictions of Hernán Casciari, based around such personages as Mirta Bertotti and even Letizia Ortiz, the blogosphere has become a masquerade ball where authors play at deception and their characters lay claim to allegedly personal areas of the web in order to explore the literary possibilities they may yield. We will analyze a number of cases in order to elucidate the influence of the perception and projection of an authorial "I" on the web with reference to the foundation of a native blog-literature, the conception of which is rooted entirely in the exploitation of these spaces in the digital world.

Keywords: Techno-aesthetics, blog-fiction, avatar, hoax, social network.

Recursos para a clasificación da produción editorial na Galiza durante a etapa franquista: deseño e alimentación da base de datos¹

CARME FERNÁNDEZ PÉREZ-SANJULIÁN
M^a ANTONIA PÉREZ RODRÍGUEZ
EDUARDO RODRÍGUEZ LÓPEZ
ÁNXELES SAAVEDRA PLACES
Universidade da Coruña
carme@udc.es

1. INTRODUCCIÓN

O proxecto multidisciplinar que presentamos é o resultado da colaboración entre especialistas en Filoloxía, Documentación, Historia e Informática; neste sentido, para alén do interese que poda suscitar polos resultados intrínsecos da propia pesquisa, representa un exemplo paradigmático de como as Humanidades Dixitais nos levan a reflexionar sobre as abordaxes metodolóxicas clásicas no campo das ciencias humanas e, ao tempo, a xerar novos modelos de investigación máis adaptados aos novos contornos dixitalizados de traballo. Estas propostas metodolóxicas permiten a descuberta e a análise eficiente de elementos culturais e patrimoniais que moi dificilmente (ou, ao menos, de xeito moito máis laborioso) poderían ser enfrontados desde a óptica clásica das Humanidades.

Así, para alén de desenvolver a metodoloxía seguida en concreto no proxecto sobre a produción editorial na Galiza durante a etapa franquista e os resultados acadados, revisaremos os erros e acertos que tivemos no curso da investigación, coa vontade de contribuír á reflexión sobre os problemas que xorden neste novo marco de investigación, na procura do desenvolvemento e/ou estandarización de novas metodoloxías de traballo no ámbito das Humanidades Dixitais.

Iniciaremos a nosa exposición cunha presentación dos obxectivos e liñas xerais do noso proxecto de investigación para, a seguir, explicarmos a metodoloxía e os criterios que se utilizaron para desenvolver o plano de traballo. No seguinte

¹ Este traballo está realizado ao abeiro de dous proxectos complementares: "La producción editorial en Galicia durante la etapa franquista (1939-1975)" (FFI2010-16924) e "Producción e mercado editorial en Galiza durante a etapa franquista: a súa incidencia na industria editorial actual" (10SEC104017PR).

apartado centrarémonos na explicación das ferramentas informáticas que se desenvolveron especificamente; no último punto analizaranse os problemas atopados no seu desenvolvemento e as solucións adoptadas para, finalmente, presentarmos as conclusións.

2. PRESENTACIÓN DO PROXECTO

O obxectivo fundamental do noso proxecto de investigación consiste no estudo da produción editorial na Galiza durante a etapa franquista. Como no resto do Estado, esta actividade veuse condicionada e limitada pola existencia dun control político que considerou como esencialmente perigosa toda a produción cultural que non respondese ao seu ideario, de modo especial a que puidese defender ou promover as daquela denominadas culturas “vernáculos” (sobre todo se se editaba en linguas diferentes ao español), xa que, neste caso, esa actividade era vista directamente como sospeitosa de separatismo.

Con todo, a pesar destas limitacións e da difícil situación económica do país, o labor editorial non foi testemuñal. Foron numerosos os proxectos editoriais que xurdiron nesta etapa na Galiza, algúns de longa duración (que chegan, mesmo, até a actualidade), outros fugaces, como pode ser o caso da colección Benito Soto, mais todos certamente importantes na historia das nosas letras. Existiron mecenas, diferentes plantexamentos editoriais e comerciais, variados modos de conseguir un público lector, e opcións estéticas e ideolóxicas diferenciadas. Contemplada no seu conxunto, toda esa actividade editorial cumpriu un papel significativo na resistencia cultural que foi chave na loita antifranquista e na procura dunha sociedade libre e democrática²; esta circunstancia non foi, na nosa opinión, suficientemente valorada até agora e, de feito, estamos a comprobar o profundo descoñecemento que existe na actualidade de moitas destas iniciativas editoriais, mesmo dalgúnhas de indiscutíbel suceso comercial³.

No campo da literatura galega, a historiografía literaria deste período consolidou a idea da centralidade cultural e político-ideolóxica dalgúnha desas editoras (debido en parte ao número de autores publicados e á incontestábel calidade literaria dalgúns dos textos alí editados). Porén, a nosa hipótese é que outros proxectos tiveron un peso moi importante na revitalización da cultura galega, nuns casos pola orixinal liña editorial desenvolvida (pénsese nas iniciativas de Xosé M^a Álvarez Blázquez), noutros polo perfil dalgún mecenas (como Isaac Díaz Pardo e Luís Seoane) ou, tamén, pola actividade levada adiante por autores que funcionaron como editores e dinamizadores culturais (como, entre outros, os poetas Manuel María ou Celso Emilio Ferreiro), sen esquecer o importante papel que tivo a autoedición para moitos dos títulos desta época e o apoio constante de imprentas e talleres de artes gráficas de moitas das vilas do noso país.

² Ao tempo, non ignoramos o extraordinario labor editorial levado a cabo nesta etapa polo exilio galego en diferentes países (sobre todo na Arxentina), nin tampouco iniciativas fundamentais para as nosas letras como as de Akal ou de Brais Pinto en Madrid. Aínda que non formen parte do noso estudo polo seu marco xeográfico, deberán ser analizados tamén nunha fase posterior.

³ Un exemplo sería o da Editorial Cíes de Vigo que, malia o seu suceso comercial, é practicamente inexistente nos andeis das bibliotecas públicas polo seu carácter de literatura “popular”.

O noso obxectivo fundamental foi recoller con vocación de exhaustividade a produción de todo tipo de iniciativas editoriais que tiveran lugar na nosa terra, aínda que parte delas non fosen literarias (velaí o exemplo da editora técnica Valle-Collantes), sen esquecermos o fundamental labor editorial de moitas das entidades públicas do momento (Academias, Universidade, museos...). Cómpre ter en conta que unha parte central da nosa investigación é o estudo das empresas que editaron en galego durante ese período (as súas características, produción, catálogos...); porén, para podermos levala adiante, só unha pormenorizada descrición do mundo da edición na Galiza daqueles anos vai permitir desenvolver con precisión a análise de moitas das circunstancias e condicionamentos en que se retoma a escrita e a edición en galego durante o franquismo na Galiza.

Na nosa opinión, neste período atópanse moitas das chaves que explican a evolución da industria editorial galega durante o posfranquismo e, aínda, en etapas máis recentes ou na actualidade. A análise en profundidade do mundo editorial, desde unha perspectiva histórica e multidisciplinar, pode ser unha fonte de ideas para o desenvolvemento de estratexias comerciais de interese no presente para este sector económico na Galiza.

3. METODOLOXÍA PARA A RECUPERACIÓN E CATALOGACIÓN DA PRODUCCIÓN EDITORIAL GALEGA (1939-1975)

A novidade do proxecto de investigación que estamos a levar adiante estriba en que pretende, por primeira vez, enfrontar a análise de todos os proxectos editoriais da Galiza da posguerra individualmente, tanto nos seus aspectos ideolóxico-culturais como comerciais, para, a seguir, comparalos entre si. Pretendemos analizar a súa relevancia na recuperación e posta en valor da cultura galega e o seu papel na resistencia cultural antifranquista durante esa etapa nos casos en que resulte pertinente.

Para cumprirmos estes obxectivos estamos a levar a cabo diversas tarefas, a máis laboriosa das cales está a ser a busca e recuperación de información da produción editorial galega entre 1939-1975, coa relación de editoriais e impresas, delimitación do lugar de edición, relación de títulos, de autores/as, etc. Este labor de fixación do catálogo resulta imprescindíbel para podermos desenvolver axeitadamente a caracterización e análise dos diferentes proxectos editoriais (equipos editoriais, coleccións, liña editorial, características específicas, tiraxes, axudas, vinculacións con outros proxectos empresariais, papel dentro do sistema literario da época, etc.), así como o estudo dos protagonistas e os seus perfís ideolóxicos (biografías, epistolarios, entrevistas...)⁴.

Neste proceso foi fundamental establecer cales ían ser os criterios a seguir na selección dos documentos (neste casos libros/obras) e dos servizos de información onde están depositados. Así, tivemos que escoller a que centros de información, bases de datos, repertorios e outras publicacións acudir para recoller o editado en Galiza entre 1939 e 1975, e que publicacións elixir.

⁴ Coa finalidade de difundir o traballo realizado, contamos con abrir ao público proximamente unha web con todos os datos e materiais recollidos. Aquí inclúiranse, para alén do catálogo e dun útil e complexo buscador, outros documentos como, por exemplo, as entrevistas e filmacións dos protagonistas aínda vivos e dos seus colaboradores directos. O enderezo será: <<http://ediciongalizafranquista.com>>.

Optouse, pois, por combinar varios procedementos; en primeiro lugar, decidimos aproveitar as vantaxes que nos ofrecen as tecnoloxías e os recursos de que hoxe dispomos na sociedade da Información; foi así como comezamos a valorar a posibilidade de recuperar as publicacións editadas neste período que custodian as bibliotecas públicas galegas (ou con fondos galegos) mediante a importación e integración de catálogos dixitalizados. Finalmente, unha vez que nos puxemos en contacto coas persoas responsábeis das bibliotecas, e tras avaliarmos as dificultades e os recursos necesarios para poder levar adiante este plano de traballo, tomouse a decisión de proceder á importación masiva de datos dos catálogos das seguintes bibliotecas públicas: Biblioteca de Galiza e Bibliotecas nodais de Coruña, Lugo, Ourense, Pontevedra, Ferrol, Vigo e Santiago.

En total, coa importación automática destes fondos documentais das bibliotecas, partimos de 22.495 rexistros bibliográficos en formato *MARC 21*. Este número de rexistros era, evidentemente, demasiado elevado e non se correspondía coa cantidade de materiais⁵ que intuíamos podían ter sido publicados naquela etapa.

Cómpre sinalar que, cando decidimos facer a importación masiva, xa eramos conscientes de que ían aparecer moitos rexistros duplicados (por estaren os exemplares repetidos en varias bibliotecas), mais contabamos con que, ao viren en formato *MARC*, sería doada a súa identificación e eliminación automática; daquela, a revisión manual ficaría só para casos problemáticos e, en principio, excepcionais. A sorpresa veu cando constatamos que os rexistros *MARC 21* dunha mesma obra presentaban numerosas diferenzas que impedían a súa identificación automática como duplicados. Esta circunstancia obrigou a un intenso traballo manual (aínda que soportado polas ferramentas informáticas desenvolvidas) de eliminación da redundancia, un labor que ocupou grande parte da actividade do grupo de investigación.

Ao tempo, fíxose a revisión e cotexo doutros materiais bibliográficos que serviron para ampliar a base de datos inicial pois, obviamente, partiamos da base de que toda a produción editorial non se conservaba nas bibliotecas⁶. Foi este un traballo laborioso que nos obrigou a revisar manualmente repertorios bibliográficos (Palau, 1948); catálogos do Instituto Nacional del Libro Español (1957-1961, 1972, 1974, 1978) e do Servicio Nacional de Información Bibliográfica (1958-168), libros e revistas editados dentro e fóra de Galiza que de forma parcial recompilan as nosas publicacións (apéndices de *Cuadernos de Estudios Gallegos*, *Galaxia Almanaque...*) ou os catálogos de diversas mostras bibliográficas feitas en distintas comarcas ou vilas galegas (Alonso y Ramos, 1963; Vilariño y Pardo, 1981; Bassave, 1989; Álvarez, 1995). Nesta fase, obviamente, resultou esencial o cotexo e revisión dos catálogos, nos casos en que se conservan, e da documentación conservada das organizacións editoriais en activo entre 1939 e 1975.

Este traballo, aínda sen rematar por completo, estase a realizar, fundamentalmente, a partir da consulta dos arquivos das propias editoriais estudadas e de bibliotecas galegas con fondos históricos (Real Academia Galega, Biblioteca

⁵ Ao falarmos de rexistros, referímonos a libros, folletos, revistas ou separatas.

⁶ Esta recuperación manual foi especialmente útil, ademais, para localizarmos moitos dos libros publicados antes do establecemento do Depósito Legal (23-XII-1957) e nos primeiros anos en que este se implantou, xa que é notorio que en todo o Estado español se publicaron bastantes libros sen el durante os primeiros anos da entrada en vigor da lei.

Xeral da Universidade de Santiago de Compostela, Instituto de Estudos Galegos Padre Sarmiento, Museo de Pontevedra, Fundación Penzol, Biblioteca da Deputación Provincial de Ourense, Biblioteca do Real Consulado da Coruña, Círculo das Artes de Lugo, Biblioteca da Fundación Manuel María...), sen deixar de lado a consulta de toda a bibliografía e documentación que se está a publicar en datas recentes sobre este tema.

4. O TRABALLO INFORMÁTICO NO PROXECTO

O traballo informático no proxecto estivo centrado en catro tarefas:

1. Construción da base de datos (BD): A partir das prioridades establecidas polos obxectivos do proxecto, deseñouse o modelo conceptual de datos, documentado nun diagrama Entidade-Relación, que recompila toda a información que se precisa na actualidade, dá cabida a ampliacións futuras e asegura a súa explotación eficiente. Este modelo foi finalmente implementado no Sistema Xestor de Bases de Datos (SXBD) PostgreSQL, un software libre que se ten consolidado nos últimos anos como unha das mellores alternativas para a implementación profesional de BD de envergadura media.
2. Desenvolvemento da aplicación web de xestión de datos: Deseñouse unha aplicación web a medida para o proxecto e implementouse con tecnoloxía J2EE. A ferramenta permite que os membros do equipo de traballo xestionen, de forma colaborativa, toda a información almacenada na BD. A aplicación dá soporte á alimentación manual e mais aos procesos de revisión; conta, tamén, cun buscador avanzado que permite construír filtros para o acceso á información almacenada combinando varios criterios.
3. Desenvolvemento das ferramentas de importación de datos: Desenvolveuse un conxunto de ferramentas que permitiron automatizar, completa ou parcialmente (dependendo do caso), o proceso de alimentación da BD a partir de fontes *MARC* (tanto *MARC21 Tagged Display*, como *MARC21 XML*). Foi preciso localizar e analizar os campos e subcampos de interese nos arquivos *MARC* e almacenalos nos campos correspondentes da BD propia.
4. Desenvolvemento da aplicación web de consulta e explotación de datos: Na actualidade, estamos a traballar nesta última tarefa de desenvolvemento da aplicación web de consulta e explotación de datos, definindo as funcionalidades que proporcionará ao público xeral e deseñando unha interface de usuario que resulte atractiva, áxil e sinxela de manexar. Esta permitirá acceder a toda a información froito do traballo de catalogación e análise realizado.

A seguir, descríbese o resultado do traballo realizado nas tres primeiras tarefas, que son as que están completadas a día de hoxe.

4.1. Construción da base de datos

Na figura 1 preséntase o Diagrama Entidade-Relación que describe a información que se recolle na BD.

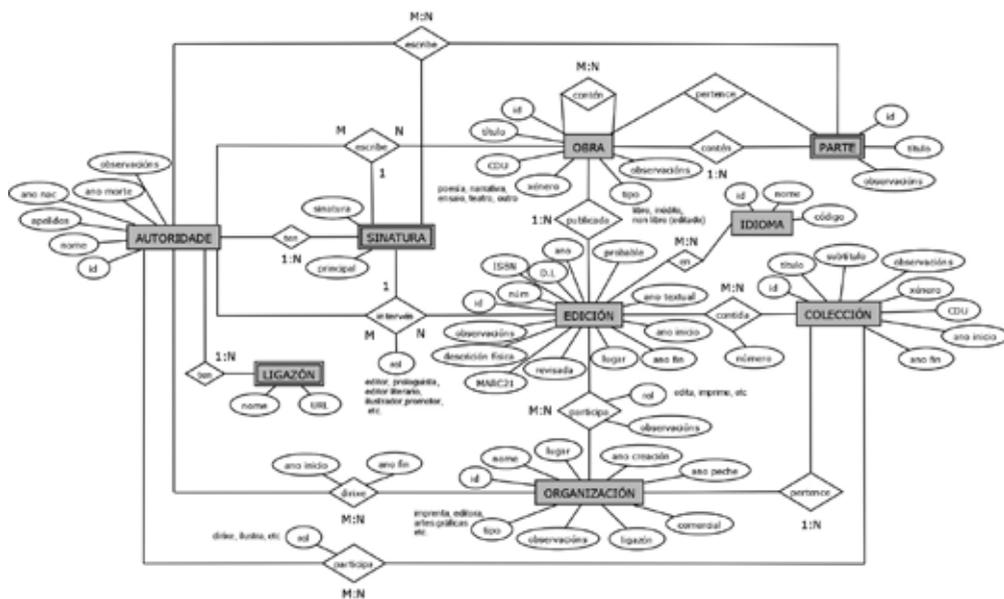


Fig. 1. Diagrama Entidade-Relación.

O centro do modelo é a entidade EDICIÓN que reflicte cada unha das “tiraxes” (edicións, reproducións, reimpresións, edicións críticas...) que se fixeron dunha obra/libro, recompilando a información específica de cada unha e deixando os datos comúns e que non varían en todas elas, co fin de os recompilar unha única vez na entidade OBRA (libro). Os atributos da entidade, que se representan cun nome dentro dunha elipse, corresponden ás características propias das edicións que se recollen na BD. Á parte, recóllese a ORGANIZACIÓN (editora, imprenta, concello, universidade...) que editou e/ou imprimiu dita edición, as/os editores, prologuistas, ilustradores, promotores, etc., isto é, as AUTORIDADES que interviñeron e a COLECCIÓN ou serie en que está esa edición, de ser o caso. Esta información descríbese a través das relacións (marcadas como rombos no diagrama) que existen entre a entidade EDICIÓN e o resto de entidades do modelo.

Do resto do modelo, cómpre destacar as estruturas que permiten almacenar información de natureza máis complexa. Referímonos, principalmente, a tres:

- a) A que permite establecer unha relación entre unha obra con calquera outra das almacenadas na BD, necesaria para contemplar, por exemplo, o caso das antoloxías.
- b) A que fai posíbel distinguir partes específicas do contido dunha obra. Esta permite, por exemplo, recuperar as diferentes edicións en que figura un poema, tanto a orixinal onde se publicou por vez primeira, como as posteriores edicións ou aquelas recompilacións en que dito poema apareza.

A que describe o feito de que algúns autores/as asinan as súas obras co seu nome real e/ou con outros pseudónimos. Permite recuperar toda a produción dun autor/a, xa se busque polo nome real ou por calquera dos seus pseudónimos.

4.2. Aplicación web de xestión de datos

A aplicación web de xestión de datos está publicada en <<http://ediciongaliciafranquista.udc.es/admin/>> mais o acceso está permitido exclusivamente ás persoas do equipo do proxecto.

Unha das interfaces deseñadas para esta aplicación é a que permite presentar un conxunto de edicións nunha listaxe tabulada (ver Fig. 2). Esta táboa deseñouse de xeito que en cada fila se amosan os datos máis representativos dunha edición: os necesarios para a permitir identificar, para acceder aos datos completos da súa catalogación (premendo no campo "Número") e mais os que permiten navegar e acceder aos datos específicos da información relacionada (a correspondente á obra, ás organizacións que editan e/ou imprimen e ás autoridades que interveñen).

Esta interface úsase na aplicación para presentar o conxunto completo de edicións da BD, ou ben un subconxunto que responde a unhas características concretas expresadas a través dos campos do buscador avanzado que ofrece a ferramenta, ou as establecidas a partir de clicks de navegación que permiten chegar ás edicións en que interveñen unha autoridade determinada, as edicións dunha colección específica, ou ás editadas/impresas por unha organización concreta.

No exemplo concreto da figura 2, pódese ver a colección dos 3.577 rexistros que existen na actualidade correspondentes a edicións de libros.

Caso de premer no "Número" da edición, preséntase a interface que permite xestionar os seus datos de catalogación (ver Fig. 3). Como se pode comprobar, para cada atributo e para cada relación da entidade EDICIÓN co resto de entida-

Fig. 2. Listaxe tabulada de edicións.

Data	Número	Obras	Autoría	Editora	Lugar	Descripción
✓ 1939	2 ^a ed. dec.	Arriba Española	Pérez de Cabo, J.	La Región	Ourense	126 p., 28 cm
✓ 1939	1 ^a ed.	Basada de a muller asturi e otras cosas	Deba, Enciso de	s.n.	A Coruña	[86] p., ...
✓ 19407	1 ^a ed.	Bando	Mosquera, F.	Monst. Imprenta: Editora	A Coruña	
✓ 1939	1 ^a ed.	Colocación obrera		s.n.	Lugo	124 p., 15 cm
✓ 1939	1 ^a ed.	Con Jasevito, modo hispico de seguir la misa	Yáñez Núñez, J. Demian	Imprenta e Papelería La Industrial	Ourense	80 p., 18 cm
✓ 1939	1 ^a ed.	Del fente de Asturias al de Madrid pasando por el quibloco (por obra de un comerciante)	López Sánchez, Enrique	La Voz de la Verdad	Lugo	147 p., 21 cm
✓ 1939	1 ^a ed.	Discursos pronunciados por D. Ramón Bermúdez de Castro y Pío Alenda Comandante Militar de Asturias en las fiestas de oratoria en dicha ciudad para solemnizar la consigna de Madrid por el ejército español y el fraustracione de la summa de recuperada nacion.	Bermúdez de Castro y Pío, Ramón	s.n.	A Coruña	65 [1] p., ...
✓ 1939	1 ^a ed.	En el centenario de la aparición de la prensa moderna del espíritu en la idea de la imprenta: discurso leído en la Academia de la Lengua Española el día 10 de Mayo de 1939.	Bullón del Castillo y Calvo de Osón, Carlos	Universidad de Santiago de Compostela	Santiago de Compostela	68 p., 27 cm
✓ 1939	1 ^a ed.	Estadístico de amor e amistad	Flórez	Biblioteca Hogaña	Lugo	79 p., 18 cm
✓ 1939	1 ^a ed.	Isaapa para el libro de Cristóbal Colón	Ray Sánchez, Luciano	s.n.	A Coruña	417 p., 22 cm
✓ 1939	1 ^a ed.	Tranco ante el Adelantado	Calvo Mauro, José	Editorial P.P.K.O.	Vigo	28 p., 28 cm
✓ 1939	1 ^a ed.	Una gran prensa de España según: nobles produccion	López García, Enrique	s.n.	A Coruña	260 p., 19 cm
✓ 1939	1 ^a ed.	Dramático del amor: curso sintético	Maggio, Luis	Biblioteca Hogaña	Lugo	72 p., 20 cm
✓ 1939	1 ^a ed.	Informe del Inspector de sanidad de la provincia de Lugo para el año de 1938	s.n.	s.n.	Lugo	82 p., 24 cm

Fig. 3. Xestión de edicións.

des do modelo, existe na interface un cadro de texto para editar a información correspondente.

Aínda que está previsto que sexa a través da aplicación pública que na actualidade se está a desenvolver desde onde se poda realmente explotar a información recollida na BD, xa a ferramenta de administración conta cun buscador avanzado (ver Fig. 4) que permite extraer datos de interese.

Así, permite realizar buscas combinando campos como o título, o ano ou o tipo de publicación, as autoridades e o papel que xogaron na edición (autor/a, ilustrador/a, editor/a, prologuista...), a organización que editou e/ou imprimiu e o idioma en que está escrita. Deste xeito, podemos coñecer xa, por exemplo, as "edicións de libros nos que un Isaac Díaz Pardo fixo de ilustrador", ou as "edicións publicadas en galego pola editorial Celta entre os anos 1968 e 1975".

A resposta da BD a unha consulta realizada a través do buscador avanzado é a listaxe de edicións que encaixan coas características de busca, que se presenta nun formato similar ao da figura 2.

Ademais do corpus de edicións, na BD contamos con tesauros relacionados de autoridades (escritores, ilustradores, prologuistas...), organizacións que editaron ou imprimiron na época e de coleccións.

A aplicación permite visualizar e xestionar os anteditos tesauros e facilita o acceso ás edicións relacionadas. Por exemplo, na figura 5, preséntase a listaxe de autoridades que temos recollidas a día de hoxe na BD. Non só podemos acceder aos datos recompilados de cada unha delas, senón que tamén temos dispoñíbel a listaxe de edicións que produciron.

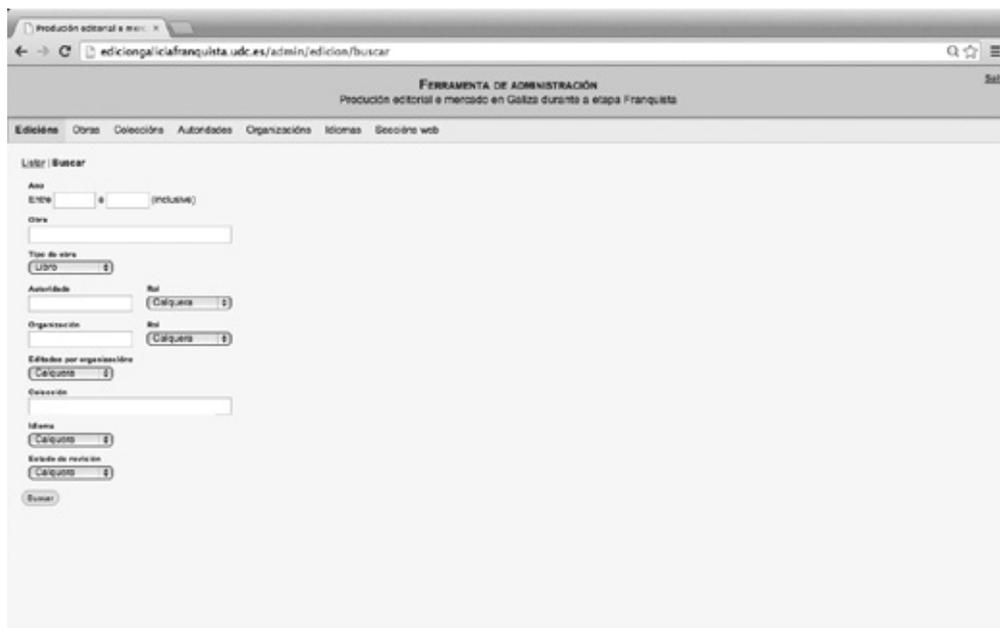


Fig. 4. Buscador avanzado.

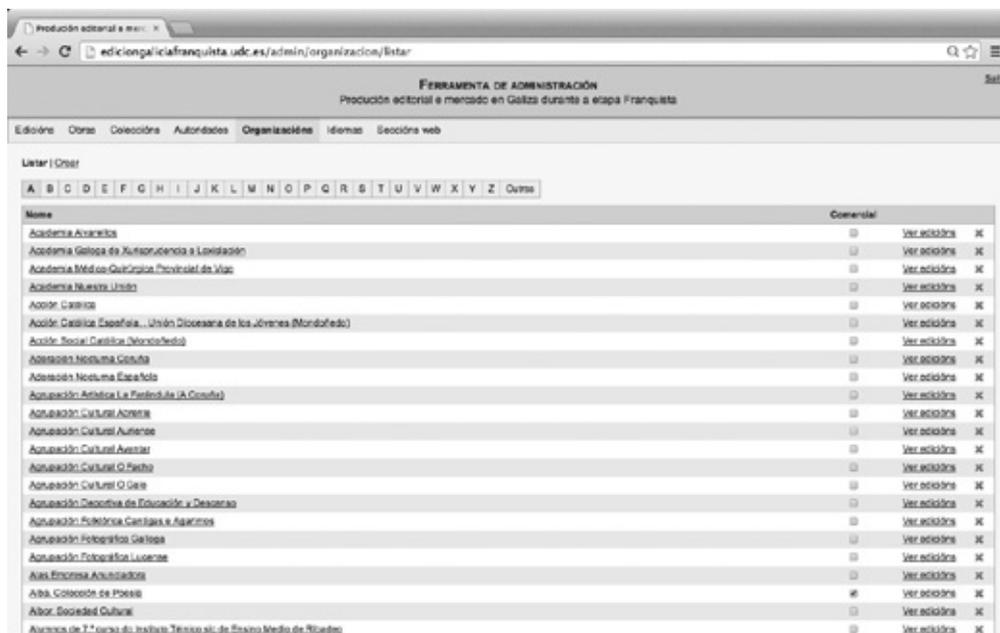


Fig. 5. Tesouro de organizacións: editoras, imprentas, etcétera.

4.3. Ferramentas de importación de datos

O obxectivo desta tarefa era importar de forma automática a información almacenada nas fontes *MARC* das que partimos, un total de 22.495 rexistros, dos cales 4.622 correspondían ao formato *MARC21 Tagged Display* (ver Fig. 6 superior) e 17.873 ao formato *MARC21 XML* (ver Fig. 6 inferior). Construíronse, polo tanto, dúas ferramentas (*parsers*), cada unha delas capaz de ler as fontes *MARC*, extraer a información dos campos relevantes para o noso contexto e almacenala nos campos correspondentes da BD.

Como xa se dixo, esperabamos que a primeira execución destas aplicacións revelase a existencia de información duplicada e, mesmo, triplicada ao pertenceren as fontes de información a diferentes bibliotecas, mais que puidese ser detectada e solucionada de forma automática, como tamén se fixo, ao afinar os *parsers* para que eliminasen signos de puntuación, espazos en branco, etc.; con todo, o maior problema veu das redundancias xeradas por culpa de non estar a información normalizada. Así, atopamos dúcias de casos en que o proceso de importación xeraba rexistros diferentes na BD para entidades ou atributos que claramente eran un só⁷. Este feito obrigou a realizar revisións exhaustivas da información das fontes *MARC*, antes e despois de as importar.

A estratexia que seguimos para evitar que o proceso de revisión fose exclusivamente manual foi extraer das fontes *MARC* listaxes de autoridades, organizacións, roles, etc. para arquivos Excel, de xeito que se puidesen detectar e resolver este tipo de redundancias. Xa con estas listaxes corrixidas, modificáronse os *parsers* con regras que permitiron almacenar as edicións asociadas ás entidades "unificadas" de forma automática.

Con todo, aínda ficou na BD moita outra información que houbo que revisar de forma completamente manual, xa que non só non era posíbel atopar patróns que permitisen construír un software que arranxase de vez todos os problemas, senón que se precisaba persoal experto que tratase e solucionase caso por caso. Porén, modificouse o esquema da BD incorporando novos atributos e desenvolvéronse funcionalidades específicas na aplicación web de xestión de datos para facilitar este traballo de revisión manual.

5. PROBLEMAS ATOPADOS E SOLUCIÓNS

Como xa foi sinalado, os maiores problemas cos que nos atopamos á hora de realizarmos o traballo de depuración e/ou revisión dos materiais importados foron os derivados da redundancia da información e da deficiente catalogación. No noso caso, esta tivo a súa orixe nos problemas preexistentes nos contidos

⁷ Isto ocorreu, por exemplo, con ORGANIZACIÓNS como "Acción Católica", para a que apareceron até 7 rexistros: "Acción Católica", "Accion Catolica", "Accion Católica", "Acción Catolica", "Acción Católica", "Acción católica; 1943]", "Acción Católica]". Tamén o detectamos con AUTORES, como "Enrique Fernández Villamil", almacenado como 5 autoridades diferentes: "Fernández Villamil y Alegre, Enrique", "Fernández Villamil, Enrique", "Fernández-Villamil Alegre, Enrique", "Fernández-Villamil y Alegre, Enrique" y "Fernndez-Villamil Alegre, Enrique". E atopamos o mesmo problema nas abreviaturas que describen os ROLES: "editor/a", "prologuista", etc., xa que variaban dun rexistro a outro con formas como "ed.", "edit." ou "edic" (para editor/a) e "pr.", "prol." ou "prolog." (para prologuista).

```

LEADER 00000nam 2200000 i 4500
001 2900515660
007 t
008 990415q1942 esp 00011 spa d
040 BUSC|bgao
090 860-1*19*
100 10 Orbe, Enrique de
245 00 Balada de la mujer estéril ; y otras poesías /|cde Enrique de Orbe ; comentadas por M. Yzquierdo y Vivas
260 1 [Galicia? :|bs.n.,|c1942?]|e(La Coruña :|fTip. El Ideal Gallego)
300 [36] p. ;|c24 cm
500 Información tomada da cuberta
700 21 Yzquierdo y Vivas, M.

```

```

<record type="Bibliographic">
  <leader>00000nam 2200000 i 4500</leader>
  <datafield tag="100" ind1="1" ind2=" " >
    <subfield code="a">Cunqueiro, Álvaro</subfield>
    <subfield code="d"> (1911-1901)</subfield>
  </datafield>
  <datafield tag="245" ind1="0" ind2="0">
    <subfield code="a">Xente de aquí e de acolá</subfield>
    <subfield code="c"> / Álvaro Cunqueiro</subfield>
  </datafield>
  <datafield tag="260" ind1=" " ind2=" " >
    <subfield code="a">Vigo</subfield>
    <subfield code="b"> : Galicia</subfield>
    <subfield code="c">, 1971</subfield>
  </datafield>
  <datafield tag="300" ind1=" " ind2=" " >
    <subfield code="a">176 p.</subfield>
    <subfield code="c"> ; 21 cm</subfield>
  </datafield>
  <datafield tag="655" ind1=" " ind2="7">
    <subfield code="a">Novela da vida cotiá</subfield>
  </datafield>
  ...
</record>

```

Fig. 6. Exemplos de Fontes MARC.

das fichas MARC 21 recibidas desde as bibliotecas que, dun modo esquemático, se poden resumir en:

- a) Falta de normalización existente nos rexistros bibliográficos dos catálogos (diferentes criterios nas denominacións, uso de abreviaturas distintas...).
- b) Erros ortográficos.
 - Duplicidades galego/español.
 - Información colocada incorrectamente nos campos MARC.
 - Ausencia de datos nas fichas MARC.

Para solucionar isto, foi preciso facer unha análise exhaustiva dos valores contidos nos distintos campos dos rexistros bibliográficos e, despois, segundo o caso, optar por diferentes tipos de procesos de depuración dos rexistros. Así, un número relativamente grande de redundancias puido ser identificado, como se indicou, de modo automático ou semiautomático mediante a comparación literal dos valores dalgúns campos.

Porén, atopamos moitos casos nos que a comparación literal de dous valores non resultou suficiente para discernir se se trataba da mesma información ou non. Por exemplo, se un rexistro contén o valor "Imprenta Moret" e outro contén "Imp. Moret", para a ferramenta informática trataríase de dúas organizacións totalmente alleas unha da outra. Dado que o período estudado é o suficiente-

mente extenso para ter que recoller mudanzas técnicas ou nominais na dirección das imprentas ou editoriais, houbo que reunificar distintas denominacións baixo unha soa: así, o que para a ferramenta informática sería unha nova organización "Gráficas do Castro-Moret", nós incluímos no mesmo rexistro "Imprenta Moret". Este mesmo problema de unificación deuse co nome de institucións (Caixas de Aforros, Deputacións Provinciais, Concellos...) que ao longo do tempo modificaron o nome e, tamén, coas variacións toponímicas que a ferramenta entendía como nomes diferentes: *Tuy/Tui*, *Pontearreas/Puentearreas*...

Outro problema consistiu na importación de rexistros que non coincidían co noso ámbito de investigación, tanto temporal (edicións posteriores a 1975, que só tiñan unha data probable de edición (197?) como xeográfica: libros con lugar de edición probable (Galiza?) ou descoñecido (s/l) que, logo de comprobar que foran editados fóra do noso marco espacial, foron borrados manualmente. Tamén se verificou e reasignou a organización editorial de boa parte dos numerosísimos rexistros que aparecían baixo a etiqueta "s.n." (sen mención editorial).

Esta exhaustiva depuración dos datos importados é a que explica a considerable redución de rexistros que se produciu desde o momento da importación dos datos ao día de hoxe (de 22.495 a 4.849 rexistros⁸). Aliás, cómpre subliñar que esta revisión só se puido facer consultando e cotexando un a un os rexistros recollidos na nosa ferramenta cos de diversas bases de datos en rede: catálogo das Bibliotecas das Universidades de Santiago de Compostela, Coruña e Vigo, catálogo da Biblioteca Nacional, do Depósito Legal, da Real Academia Galega, Biblioteca Dixital de Galicia, etc. ou, mesmo, co cotexo físico dos libros. Esta fase está a ser de enorme interese pois nos permite completar datos que non aparecen na ficha *MARC 21* da que partimos, como a información referida a coleccións ou series (que en moi poucos casos aparecían recollidas nas fichas catalográficas da época), así como outros outros datos complementares de grande interese (autoría de limiares, prólogos, capas, ilustracións...).

Ao longo deste proceso de revisión constatamos tamén a necesidade de crear as fichas de certas obras que, constándonos a súa existencia, non aparecían na importación dos datos das bibliotecas (isto é, non existía ningún exemplar das mesmas nas bibliotecas públicas galegas). Estas incorporacións manuais ao programa informático non levan incluída, como é lóxico, a ficha *MARC 21*.

Aínda que unha BD destas características sobre a edición en Galiza de 1939 a 1975 é un traballo que dificilmente estará de todo rematado (pensemos nas innumerábeis edicións de autor que aparecen en calquera biblioteca particular que revisemos), sabemos que a nosa recompilación de datos ofrece unha nova perspectiva sobre o panorama editorial da época até o de agora nunca enfrontado no seu conxunto.

CONCLUSIÓNS

A partir da nosa experiencia neste proxecto de investigación, podemos extraer dous tipos de conclusións, unhas relacionadas cos aspectos metodolóxicos e outras cos contidos do noso traballo.

⁸ Aínda que, como xa se explicou, este número de rexistros inclúe libros, folletos, separatas e revistas, o noso traballo centrábase nos 3.300 títulos de libros que temos actualmente na base de datos.

Canto á metodoloxía, primeiro, parécenos imprescindible insistir en que a colaboración entre equipos interdisciplinares de Humanidades e de Informática é básica para poder afrontar estudos como o desenvolvido neste proxecto. En segundo lugar, aínda que a día de hoxe a dixitalización chegou a todas partes (sen ir máis lonxe, aos catálogos das bibliotecas) e que, por tanto, se pode facer un tratamento informático deses datos, a realidade é que estamos nun tempo de transición onde, por máis que teñamos ferramentas informáticas e catálogos dixitalizados, as metodoloxías de traballo son aínda, por así dicilo, "pre-informáticas". Faltan, pois, normalización, estandarización, protocolos de traballo, procedementos... liñas dun camiño que haberá que ir facendo.

Finalmente, canto aos resultados do noso proxecto, podemos afirmar que o nosa traballo de recompilación fornece unha visión novedosa do panorama editorial na Galiza da época franquista, cuxa análise en conxunto non fora feita até agora. Unha vez rematada, esta será unha Base de Datos de referencia para o estudo deste período.

BIBLIOGRAFÍA

- Alonso Montero, Xesús y Ramos de Castro, Epifanio, *Cen anos de literatura galega. Catálogo da exposición bibliográfica 9/XII/1963-29/XIII/1963*, Lugo, Círculo das Artes, 1963.
- Alvarez García, Isabel, *O Libro Galego, onte e hoxe 1993: catálogo da Exposición bibliográfica, Santiago, xaneiro 1995*, Santiago, Federación de Libreiros de Galicia, 1995.
- Bassave Roibal, Rosa M^a et al., *O Libro Galego, onte e hoxe 1988: catálogo da Exposición bibliográfica, Santiago, maio 1988*, Santiago, Federación de Libreiros de Galicia, 1989.
- Cuadernos de Estudios Gallegos*, apéndice Bibliografía de Galicia, tomos I-XI (1945-1956), XIII (1958), XIX-XXIII (1964-1968).
- Galaxia Almanaque: 1950-1975*, Vigo, Galaxia, 1975.
- Instituto Nacional del Libro Español (ed.), *Catálogo General de la Librería Española 1931-1950*, Madrid, Instituto Nacional del Libro Español, 1957-1961.
- Instituto Nacional del Libro Español (ed.), *Índice de la producción editorial española 1968-1972*, Madrid, Instituto Nacional del Libro Español, 1972.
- Instituto Nacional del Libro Español (ed.), *Catálogo de libros en gallego*, Madrid, Instituto Nacional del Libro Español, 1974.
- Instituto Nacional del Libro Español (ed.), *Catálogo de libros en gallego*, Madrid, Instituto Nacional del Libro Español, 1978.
- Palau y Dulcet, Antonio, *Manual del librero hispanoamericano: bibliografía general española e hispano-americana desde la invención de la imprenta hasta nuestros tiempos con el valor comercial de los impresos descritos*, Barcelona, Antonio Palau, 1948.
- Servicio Nacional de Información Bibliográfica (ed.), *Bibliografía Española*, Madrid, Dirección General de Archivos y Bibliotecas, 1958-1968.
- Vilariño Pintos, Daría y Pardo Gómez, M^a Virtudes, *O Libro galego onte e hoxe. Catálogo da exposición bibliográfica, Santiago, maio-xuño 1979*, Santiago, Secretariado de Publicacións da USC, 1981.



RESUMO

Explicación da metodoloxía e das ferramentas informáticas desenvolvidas para acometer un proxecto de investigación propio das Humanidades dixitais. En concreto, este proxecto ten como obxectivo fundamental estudar a produción editorial na Galiza durante

a etapa franquista (1939-1975), analizar se existiu unha industria cultural editora e en que medida esta funcionou como canle difusora, tanto dunha liña de resistencia política contra o franquismo, como noutra de aposta pola conservación e posta en valor da identidade cultural da Galiza.

Tras unha breve introdución ao proxecto, preséntase a metodoloxía seguida no proceso de recuperación e catalogación dos fondos, explícase como se desenvolveu o proceso de importación masiva dos catálogos das bibliotecas públicas máis relevantes da Galiza, así como as súas características técnicas e, a seguir, descríbense as ferramentas informáticas deseñadas. Finalmente, sinálanse os problemas metodolóxicos atopados e as solucións que se foron adoptando pois considérase que poden ser de interese para outros proxectos do mesmo tipo. Coidamos que o noso modo de enfrontalos e a experiencia obtida á hora de resolvelos constitúen, de seu, achegas relevantes deste proxecto ao mundo das Humanidades dixitais.

Palabras clave: Producción editorial, base de datos, clasificación bibliográfica, franquismo, Galiza.

ABSTRACT

Explanation of the methodology and the computational tools developed to undertake a research project within the domain of the Digital Humanities. In particular, the main goal of this project is to study the publishing industry in Galicia during the Franquist period (1939-1975), to analyze whether there was a culturally-concerned publishing industry and to determine the extent to which this acted as a means of public diffusion, both of political resistance against Franquism as well as the commitment to the preservation and enhancement of the Galician cultural identity.

After a brief introduction to the project, we present the methodology followed throughout the process of retrieval and cataloging of the lists of titles, we explain now the process of the massive importation of the catalogs of the most important Galician public libraries was carried out, as well as their technical features and, subsequently, we describe the computational tools designed. Finally, we point out the methodological problems found and the solutions adopted given their possible usefulness for similar projects. We consider that the way they were dealt with and the experience gained solving them constitute, in themselves, important contributions of this project to the area of digital humanities.

Keywords: Publisher production, Data bases, Bibliographic compilation, Franquism, Galiza.

Base de datos sobre Inventarios y Bibliotecas del Siglo de Oro (IBSO). Utilidad y posibilidades

CARLOTA FERNÁNDEZ TRAVIESO
Universidade da Coruña
cfernandez@udc.es

PRESENTACIÓN Y FINALIDAD DE IBSO

Desde el año 2007 el grupo SIELAE ha trabajado en la creación y alimentación de una compleja base de datos que se ha dado a conocer con el nombre de IBSO, acrónimo de "Inventarios y bibliotecas del Siglo de Oro". Esta aplicación informática reúne y cruza una serie de información que habitualmente hallamos dispersa:

- Listas de libros mencionados en inventarios de bibliotecas que existieron en el Siglo de Oro y, de ser posible, las identificaciones de estas menciones con una obra, una edición y un ejemplar concretos.
- Obras de autores de este período y las ediciones de sus obras.
- Las lecturas mencionadas en obras del siglo XVI o XVII que se recomendaban para adquirir pericia en ciertos temas (en letras humanas, en oratoria o en pintura artística), lo que en IBSO se denomina "Bibliotecas ideales".
- Las fuentes que un determinado autor cita con detalle en sus propios textos y las que sabemos que manejó porque dejó rastros físicos (como su firma o notas) en ejemplares que conservamos, lo que en IBSO se conoce como "Bibliotecas hipotéticas".
- Descripciones de obras, ediciones y ejemplares.

La finalidad de hacer acopio de toda esta información es doble. Por una parte, poner toda esa información al alcance de los investigadores a través de la red de redes en un único sitio web. IBSO permitirá que ahorremos mucho tiempo al poder consultar muchos datos sin movernos de nuestro escritorio y sin necesidad de acudir a diversas fuentes. Por otra parte, gracias a la posibilidad de cruzar esa información que ofrece esta base de datos relacional, podremos, por ejemplo, obtener indicios de la popularidad de una obra dada la presencia de ésta en varios inventarios de bibliotecas del Siglo de Oro; saber fácilmente cuántas ediciones hay de una obra de un autor, incluyendo aquellas que se publicaron bajo un título global que no corresponde con el que nuestro investigador busca (me refiero, por ejemplo, a ediciones de obras que se incluyen en un vo-

lumen de obras completas de un autor); podremos saber también qué obras recomendadas en una biblioteca ideal determinada han sido citadas por un autor concreto; en qué obras un autor determinado cita a otro autor en particular; qué autoridades han sido citadas por más de un número X de escritores; cuántas obras de una determinada materia se encontraban en las bibliotecas del Siglo de Oro que conocemos; etcétera.

Toda esta información y distintas posibilidades de búsqueda se ofrecen gratuitamente en la dirección de Internet <<http://www.bidiso.es/IBSO>>. En esta interfaz de consulta se accede a la información a través de cinco vías: Onomástica, Ediciones, Inventarios, Bibliotecas ideales y Bibliotecas hipotéticas. Inventarios es la vía en la que se ha producido un mayor avance: hay 16 inventarios completos y 3 en los que estamos trabajando, con 27.154 registros introducidos, muchos de los cuales identificados. Pueden consultarse inventarios tan importantes como el de Lorenzo Ramírez de Prado; el de Diego Sarmiento de Acuña (conde de Gondomar); el de Vicencio Juan de Lastanosa; el de Antonio Juan Luis de la Cerda (VII duque de Medinaceli), quien fue buen amigo de Quevedo, o el de la biblioteca del Monasterio de San Martín realizado en 1788. Contamos además con pantallas para Búsqueda simple, Búsqueda avanzada y Otras búsquedas, en la que se ofrecen una serie de pesquisas parametrizadas. Para más detalles sobre el funcionamiento de esta interfaz de consulta y la finalidad de esta base de datos, se puede consultar un artículo publicado en la revista *Etiópicas* al que un enlace en la web de IBSO da acceso (<http://www.uhu.es/revista.etiopicas/num/07/art_7_1.pdf>).

Además de esta interfaz de consulta, se ha diseñado también una interfaz de administración a través de la cual los miembros del equipo y quienes deseen colaborar con nosotros pueden consignar los datos. IBSO es una aplicación muy ambiciosa que precisa del trabajo colaborativo para cumplir en mayor grado y en menos tiempo con sus metas, por eso se ha pensado en la posibilidad de que investigadores externos puedan acceder a la interfaz de administración con sus propias claves. Para obtenerlas, quienes deseen sumar sus esfuerzos con los nuestros han de ponerse en contacto con la dirección del equipo a través del correo electrónico <emblemco@udc.es>.

TAREAS DESARROLLADAS EN LA CREACIÓN DE IBSO

En la creación de IBSO ha sido imprescindible la colaboración entre los miembros del grupo SIELAE (expertos en filología y bibliografía) y los miembros del laboratorio de bases de datos de la Universidade da Coruña. Hemos trabajado juntos para modelar los datos y obtener el diagrama Entidad-Relación que, posteriormente, permitió a los informáticos implementar la base de datos con MySLQ (el sistema gestor de datos elegido) (Fig. 1). En una primera fase, nuestro papel consistió en:

- Analizar las fuentes de las que extraeríamos la información que volcaríamos en IBSO (por ejemplo, inventarios recogidos en trabajos como *Libros lectores y lecturas. Estudios sobre bibliotecas particulares del Siglo de Oro* de Trevor J. Dadson (1998) en el que se nos proponía ya gran parte de las identificaciones de los registros o en manuscritos como el de la Biblioteca del Monasterio de San Martín de 1788; obras de los siglos XVI y XVII como el *Discurso de las letras humanas* de Baltasar de Céspedes (Andrés, 1965)

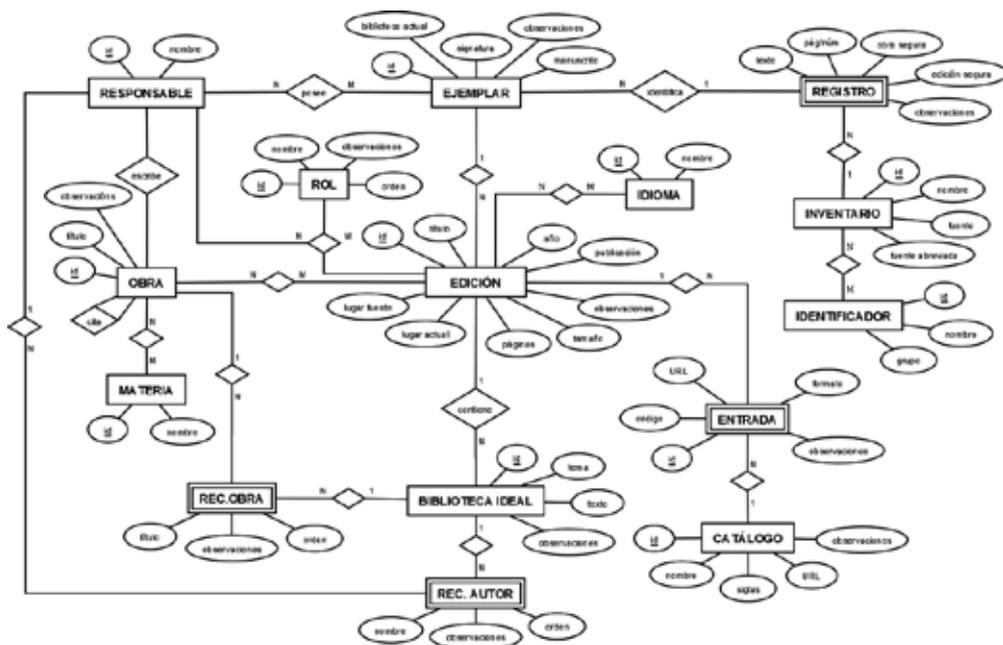


Fig. 1. Diagrama E-R de IBSO.

o el *Museo pictórico y escala óptica* de Acisclo Antonio Palomino de Castro y Velasco (1715) en las que se nos dan recomendaciones de lecturas).

- Pensar en qué información de interés podríamos aportar como expertas en filología para completarlos (por ejemplo, identificaciones de los registros de inventarios con una obra, una edición y de ser posible con un ejemplar concreto, cuando no han sido aportadas por otros investigadores o descripciones de ediciones y ejemplares).
- Reflexionar a qué preguntas y necesidades queríamos dar respuesta cruzando toda esta información.

Las conclusiones de nuestro trabajo se unieron al conocimiento de los expertos en creación de bases de datos para, juntos, determinar:

- Sobre qué elementos queríamos guardar información, es decir qué "Entidades" debía tener nuestro diagrama: las representadas en el diagrama con un rectángulo (Inventario, Registro, Ejemplar, Edición, Obra, Responsable, Biblioteca ideal, Recomendación de autor, Recomendación de obra, etcétera).
- Qué características íbamos a describir para cada una de ellas, es decir los "atributos" que teníamos que hacer constar en el diagrama: los representados con una elipse (por ejemplo, título, editor, lugar, año, tamaño, páginas, etc. para la entidad Edición).
- Cómo se relacionaban entre sí las distintas Entidades y qué entidades eran comunes entre Inventarios, Obras y ediciones, Bibliotecas ideales y Bibliotecas hipotéticas para enlazar así todos los datos, esto es dilucidar cómo

se estructura la información. Estas relaciones se representan en nuestro diagrama con un rombo. En el diagrama se recoge, por ejemplo, que un responsable escribe una obra, una obra tiene ediciones y las ediciones tienen ejemplares o que un inventario tiene registros, estos registros se identifican con ejemplares, ediciones y obras. Es importante ver que estas vías confluyen: todos los ejemplares se guardan juntos, todas las ediciones juntas, todas las obras juntas independientemente de la vía por la que introduzcamos la información. Esto es imprescindible para que la base de datos funcione).

- Además, para cada entidad vinculada a otra determinamos el número de entidades que participan de esa relación, la cardinalidad, que se representa con letras mayúsculas y números junto a cada una de las relaciones. Así por ejemplo, gracias a esta representación sabemos que cada inventario tiene varios registros, de ahí que se haga constar una N junto a los registros y cada registro pertenece a un solo inventario, se ahí que se haga constar un 1 junto a los inventarios. Asimismo, sabemos que cada obra puede tener más de una materia y cada materia puede ser asignada a más de una obra, por eso se hace constar una M y una N para esta relación.

En una segunda fase, una vez implementada por los informáticos una primera versión de la base de datos, tuvimos que probar a introducir la información y documentar cuáles eran los problemas al hacerlo si los había, para, de nuevo juntos, modificar adecuadamente el diagrama. Así llegamos a obtener nuestro diagrama Entidad-Relación (Fig. 1), el que hasta hoy consideramos definitivo.

En la actualidad, en una tercera fase de nuestro trabajo, estamos llevando a cabo la alimentación de la base de datos, en la que como antes apuntamos la colaboración de otros investigadores es bienvenida.

DIFICULTADES SURGIDAS EN LA CREACIÓN DE IBSO Y SOLUCIONES PROPUESTAS

Al elaborar y alimentar la base de datos nos hemos encontrado con un numerosos retos y problemas. A partir de ahora me centraré en ellos y en explicar cómo les hemos dado solución.

Al elaborar la base de datos

Hemos encontrado dificultades al tratar de discernir qué entidades eran necesarias para poder plasmar toda la información que queríamos volcar en IBSO. Por ejemplo, inicialmente omitimos la entidad Registro, de la cual procede el campo de la base de datos en el que hacemos constar por completo cada una de las descripciones de los ejemplares recogidas en un inventario. En la primera versión de IBSO, la entidad Inventario se vinculaba directamente a Ejemplar. Así, en el campo Ejemplar de la base de datos introducíamos la parte de texto del inventario en la que se aludía a la obra. Por ejemplo, al encontrar la descripción "*Abad Carrillo. Anales y memorias cronológicas primera y segunda parte. Huesca 1622.*" del inventario de la biblioteca de Vicencio Juan de Lastanosa, en el campo Ejemplar, transcribíamos "Anales y memorias cronológicas primera y segunda parte". Reservábamos la información sobre el autor ("Abad Carrillo") o la edición ("Huesca 1622") como ayuda para identificarlos. El nombre del autor

uniformizado (Carrillo, Martín [1561-1630]) y los datos de la edición que se completaba con ayuda de catálogos (*Annales y memorias cronológicas. Contienen las cosas mas notables assi ecclesiasticas como seculares succedidas en el mundo señaladamente en España desde su principio y poblacion hasta el año M.DC.XX*, Huesca, Pedro Bluson, Viuda de Juan Perez Valdiuielso, 1622, CCPB000039931-0) se consignaban en las pantallas y campos oportunos para que pudieran a su vez relacionarse con otros ejemplares de otros inventarios, pues podría haber más obras de ese autor y más ejemplares de esa edición en otras bibliotecas. De aquí surgieron múltiples problemas:

- Si el inventario era muy impreciso describiendo los ejemplares (por ejemplo, en el inventario de Diego Hurtado de Mendoza nos encontramos con esta alusión: "*Vn quaderno cubierto de pergamino y escripto de mano en lengua valencyana*"), no sabíamos si introducir en el campo Ejemplar información de naturaleza tan distinta al ejemplo anterior, en el que se mencionaba el título de la obra, u omitir estos datos.
- Si no conseguíamos identificar el autor o la edición, los datos que el inventario aportaba se perdían, lo que dificultaba que, posteriormente, otros investigadores intentaran la identificación.
- Si teníamos la fortuna de llegar a descubrir que el ejemplar descrito en el inventario se hallaba en la actualidad en alguna biblioteca, no teníamos lugar para hacer constar datos para su localización (biblioteca y signatura).
- Si un mismo ejemplar hubiera estado en dos bibliotecas (por ejemplo en las bibliotecas del primer duque de Pastrana y del tercero, su descendiente), no hubiéramos tenido modo de hacerlo constar.

Así pues, fue imprescindible añadir la entidad Registro. No habíamos considerado que si bien los inventarios apuntan a ejemplares, los inventarios contienen sólo menciones a ellos, que pueden ser más o menos precisas, completas o erróneas, pero que en cualquier caso no es información desdeñable, pues puede llevarnos a la identificación con una de una obra y su autor, una edición y, eventualmente, el ejemplar en la actualidad. De manera similar omitimos la entidad Recomendación en las Bibliotecas ideales. Éstas se vinculaban directamente a obras. No habíamos considerado que en muchos casos no seríamos capaces de identificarlas y que, si no incluíamos este campo, perderíamos mucha información: el fragmento de texto de estas obras en el que se mencionan las obras recomendadas, que quizás otros investigadores podrían llegar identificar.

También hemos tenido dificultades a la hora de establecer las relaciones entre las entidades. Así por ejemplo, inicialmente las recomendaciones de las Bibliotecas ideales se podían identificar sólo con una obra, pero encontramos que en muchos casos se recomendaba también la lectura de un autor en general. Se solucionó creando la entidad Recomendación de autores que se vinculaba directamente a la entidad Responsable, sin necesidad de pasar por la entidad Obras. Otro problema que surgió es que, inicialmente, en lugar de la entidad Responsables teníamos la entidad Autores. En ella introducíamos además de a autores, a traductores o comentaristas, pero excluíamos a impresores y editores, para los que existían los campos oportunos en la pantalla de datos editoriales. Sin embargo, más allá de una cuestión de nomenclatura o de concepción de la entidad el problema residía en que esta entidad autores se vinculaba únicamente a la entidad Obra. De este modo la información no quedaba estructurada correc-

tamente, pues traductores, comentaristas o adicionadores lo son normalmente de ediciones concretas y no de todas las ediciones de una obra. Además, de esta manera, se podía perder información o se duplicaba: la única manera de constatar que una persona introducida en la tabla autores era el traductor era ponerlo entre paréntesis detrás de su nombre; si una misma persona había sido traductor en algunos casos y autor en otros debía introducirse varias veces y si no se consignaba la función que este había tenido, se perdía información. Asimismo, también se disgregaría la información si surgían casos en los que un editor o impresor hubiera sido, por ejemplo, también el autor o el comentarista del mismo (no es inusual, por ejemplo, que el autor de una Relación de sucesos sea también su editor). Para solucionar estos casos hemos creado la entidad Responsables, que se concibe como un índice onomástico que reúne los nombres de todas las personas introducidos en IBSO. Esta entidad se vincula a Obra y a Edición, para permitir que los autores se puedan vincular a una obra como sus responsables y los demás responsables (impresores, editores, traductores, comentaristas, etc.) a una edición. Además se creó la entidad función, para poder precisar, eligiendo de un desplegable del campo edición, qué papel tuvo la persona en cuestión.

Igualmente, hemos encontrado dificultades al determinar la cardinalidad de las relaciones. Por ejemplo, inicialmente, habíamos considerado que de cada obra podía haber varias ediciones y que cada una de estas ediciones lo eran de una obra en particular. Sin embargo, nos encontramos con ediciones que contenían más de una obra. En el inventario de la biblioteca de Juan Vicencio Lastanosa nos encontramos con el siguiente caso: un registro, "Cayo Crispio Salustio sus obras Madrid 1632 8º", que apuntaba a un ejemplar, hoy en paradero desconocido, de la edición de las obras de Salustio traducidas por Enmanuel Sueiro, publicadas en Madrid en 1632. Este volumen contenía el *Bellum Jugurthinum*, *De Catilinae Coniuratione* y *Orationes in Catilinam*. En la primera versión de la base de datos no había modo de reflejar este caso, pues la cardinalidad entre las entidades obras y ejemplar debía ser M:N y no 1:N. También hubo que cambiar la cardinalidad para recoger el hecho de que algunas obras pueden tener más de un autor o responsable y que una edición puede estar en más de un idioma.

Sin embargo, el mayor de los retos fue el hallar sistemas para evitar duplicar la información. Aunque consiguiéramos una correcta estructuración de la información que permitiera que los datos coincidentes confluyeran (lo que resulta imprescindible para que la BD funcione), en IBSO, los materiales se introducen por diversas vías y distintas personas. Fue imprescindible instalar buscadores en muchas de las pantallas de la interfaz de administración y establecer un procedimiento por el cual, antes de introducir una nueva información (un nuevo onomástico, una nueva obra, una nueva edición, una nueva materia...) hay que comprobar si ya está en la BD, pues la información nueva debe vincularse a los datos ya existentes y sólo han de crearse un onomástico, una obra, una edición o una materia en caso de que no existan, para evitar la disgregación de datos. Así mismo, de algunos nombres de persona existen variantes y, a veces, podemos referirnos a una misma obra con otro título; sin embargo, debemos fijarnos en cuál es la forma por la que se ha optado anteriormente en IBSO y utilizarla. Por ejemplo, Andrea Alciati es mencionado en las distintas ediciones de sus obras de distintas maneras: André Alciat, Andrea Alciato o Andreae Alciati. Para evitar introducir a este autor 4 veces en IBSO y conseguir que todas sus obras aparezcan unidas bajo su nombre, hemos fijado una convención para ele-

gir sólo una de las posibles variantes de su nombre. Para la obra *Emblematum liber* de Andrea Alciato, se documentan diferentes títulos de edición *Emblematum liber*, *Emblematum libellus*, *Emblemata*, *Los Emblemas de Alciato [...]*, *Livret des emblems [...]*, etcétera. Para poder introducir cada uno de los títulos y al mismo tiempo hacer constar que se trata de una misma obra tuvimos que diferenciar entre el Título normalizado (una forma acordada del título de una obra, bajo el que reunimos todas las ediciones) y título de la edición, que puede ser distinto para cada una de ellas.

Al consignar los datos

Identificar los registros de inventarios es uno de los mayores retos que se presentan al alimentar IBSO. Aunque algunos inventarios como el de la Biblioteca del conde de Gondomar o Lastanosa son más precisos, muchos de ellos se refieren a los ejemplares que la biblioteca contenía con gran vaguedad. Por ejemplo, en el inventario de 1632 de la biblioteca de Diego Hurtado de Mendoza, I conde de Mélito, el registro número 32 dice: "Vnas ¿horas? Enbueeltas en pergamino" que, según afirma Trevor J. Dadson, es imposible identificar. Con todo, en muchos casos se puede obtener una identificación parcial, posible o limitada a dos o tres opciones. Cuando esta tarea no ha sido realizada por otro investigador, nos aventuramos a completar la información. La tarea de identificar implica: primero suponer a que autor y obra se refiere y realizar las búsquedas oportunas para recoger el nombre completo y más correcto del autor y su obra, comprobar si ese autor escribió el texto en cuestión, etcétera; a continuación realizar búsquedas en catálogos (algunos de ellos informatizados como el Catálogo Colectivo del Patrimonio Bibliográfico Español) hasta dar con la posible edición y eventualmente, aunque esto es muy difícil, a menos que se halle un *ex libris*, el ejemplar concreto, si aún se conserva en la actualidad. Para ayudar al investigador a plasmar los resultados de sus pesquisas IBSO permite:

- Realizar una identificación gradual (sólo de la obra o sólo de la obra y la edición, sin necesidad de comenzar por el ejemplar).
- Consignar parcialmente datos de la obra (por ejemplo, si hemos identificado el autor, pero no el título) y de la edición (por ejemplo, si conocemos el traductor, pero no tenemos más datos de la edición).
- Precisar si una información es sólo posible, marcando una casilla obra/edición no identificada con seguridad.
- Y relacionar una obra con la que hemos identificado un registro con más de una edición posible, ya que limitar el número de opciones a dos o tres supone ya un avance.

También supone gran dificultad el decidir la materia (o materias) a la que adscribir una obra. No sólo porque en muchas ocasiones las obras que registramos sean desconocidas para nosotras y debamos buscar información sobre ellas, sino porque no hemos hallado ningún listado de materias que sea adecuado para clasificar las obras del Siglo de Oro. Muchas de las materias que entonces se consideraban como tales hoy han desaparecido a la hora de ordenar las bibliotecas. Por ejemplo, nos hemos hallado con la duda de si diferenciar los libros de materias como la Historia Sagrada de los libros religiosos, pues en la época se consideraba que ésta era parte de la Historia. No sabíamos bajo qué marbete

clasificar obras que hoy consideramos parte de la narrativa, pues aunque en la época existían libros de poesía y de comedia, no se consideraba la narrativa como materia. Los libros de caballerías, por ejemplo, se solían clasificar bajo el marbete "libros fabulosos", una categoría que valoramos si incluir o no. Además, existían materias que hoy no tendríamos en consideración salvo en una clasificación muy detallada o especializada, este es el caso de la albeitería, que se diferenciaba debido a la importancia de los caballos en la época.

Así mismo, para otros casos que no hayan sido pensados y que se puedan presentar eventualmente se ha introducido un campo Observaciones en varias pantallas, como por ejemplo la de identificación.

CONCLUSIÓN

A pesar de estas dificultades al alimentar la base datos, en la medida en que los trabajos vayan avanzando, IBSO se convertirá en una herramienta de gran ayuda para el avance del conocimiento en la historia de la lectura, la historia de la imprenta, el uso de las fuentes, la trasmisión del conocimiento o la historia editorial de una obra. Un avance que será posible gracias a la alianza de las humanidades con la informática en la Universidade da Coruña.

BIBLIOGRAFÍA

- Andrés, Gregorio de, *El Maestro Baltasar de Céspedes, humanista Salmantino y su discurso de las letras humanas. Estudio biográfico y edición crítica*, El Escorial, Biblioteca "La ciudad de Dios", 1965.
- Dadson, Trevor J., *Libros, lectores y lecturas. Estudios sobre bibliotecas particulares del Siglo de Oro*, Madrid, Arco Libros, 1998.
- Índice general de la biblioteca del real y parroquial monasterio de San Martín de Madrid. Se hizo siendo abad el R.P.M.F. Plácido Vicente. Año MDCCLXXXVIII, Real Academia de la Historia, Ms. 9-2099.
- Palomino de Castro y Velasco, Antonio, *El museo pictórico, y escala óptica. Tomo I. Teórica de la pintura en que se describe su origen y se prueban, con demostraciones matemáticas y filosóficas sus más radicales fundamentos*, Madrid, Lucas Antonio de Bedmar, 1715.



RESUMEN

IBSO, la compleja base de datos relacional sobre inventarios y bibliotecas de los siglos XVI y XVII creada por el grupo SIELAE, reúne y pone al alcance de los investigadores, de modo gratuito a través de Internet, el resultado de varios tipos de trabajos habitualmente dispersos. Cruzando la información de sus cuatro secciones (Inventarios, Obras y ediciones de autores del Siglo de Oro, Bibliotecas ideales y Bibliotecas hipotéticas), esta aplicación permite dar respuesta a muchos de los interrogantes y necesidades de los investigadores sobre cultura europea de la Edad Moderna o de quienes realizan trabajos de edición filológica de obras del Siglo de Oro, la historia de la lectura, el uso de fuentes, la historia de la imprenta, la transmisión del conocimiento o la historia editorial de un texto.

Este trabajo ofrece información sobre la finalidad y posibilidades de búsqueda de IBSO. Además, se detalla qué tareas han sido llevadas a cabo para su creación (desde

su diseño conceptual hasta su alimentación), se muestra el diagrama Entidad-Relación de esta base de datos, se describen las dificultades y problemas que han surgido al elaborar la base de datos y al proceder a consignar los datos y se explican las soluciones a las que hemos llegado.

Palabras clave: Base de datos, inventarios, bibliotecas, siglo XVI, siglo XVII, erudición, lecturas, fuentes, ediciones.

ABSTRACT

IBSO, the complex relational database on inventories and libraries of the sixteenth and seventeenth centuries created by the SIELAE group, gathers and makes available to researchers the result of several types of jobs usually scattered, in open access via the Internet. This application allows to answer many of the questions and needs of researchers on European culture of the modern age, as well as who produce works of philological edition of Golden Age books, the history of reading, the use of sources, the history of printing, the knowledge transmission or the history of a publication. It achieves this result by crossing the information introduced into its four sections: Inventories, Works and editions of Spanish Golden Age authors, Hypothetical libraries and Ideal libraries.

This work provides information on the purpose and search possibilities of IBSO. Besides, the work details the tasks that have been carried out for the database creation (from conceptual design to feeding), shows the Entity-Relationship model of this database, describes the difficulties and problems founded while developing the database and while proceeding to record the data, and explains the solutions we have reached.

Keywords: Database, inventories, libraries, XVI century, XVII century, erudition, readings, sources, editions.

El canon literario hispanoamericano en la era digital

CAROLINA FERRER
Université du Québec à Montréal
ferrer.carolina@uqam.ca

El concepto de canon es uno muy difícil de definir. Según algunos textos de referencia (Guillory, 1995; Lentricchia y McLaughlin, 1995), etimológicamente, la palabra proviene del griego *kanon* que significa regla o vara para medir. Inicialmente, el término es empleado para identificar aquellos libros del Antiguo y del Nuevo Testamentos que cuentan con la aprobación de las autoridades eclesiásticas en cuanto corresponden a la palabra divina, es decir aquellos textos que constituyen las Sagradas Escrituras. En las últimas décadas, en los estudios literarios se considera que la canonización de los textos literarios opera de manera análoga al proceso de constitución del canon bíblico.

En 1995, al comienzo de su libro *El canon occidental: los libros y la escuela de los años*, el crítico estadounidense Harold Bloom determina que:

Originalmente el Canon significaba la selección de libros en nuestras instituciones de enseñanza y, a pesar de las recientes políticas de multiculturalismo, la verdadera pregunta del Canon sigue siendo: ¿qué debería tratar de leer un individuo que desea leer a estas alturas de la historia? (1995: 15)¹.

Vista así, la pregunta fundamental del canon parece simple. Sin embargo, se trata de una noción significativamente compleja que ha suscitado polémicas considerables. Así, por ejemplo, en 1983, *Critical Inquiry* publica un volumen completo dedicado al concepto del canon. En su introducción, Robert von Hallberg establece que:

El interés en los cánones forma sin duda parte de una investigación más amplia dentro de las instituciones de estudio literario y de producción artística. 'Política', 'economía', 'social', 'autoridad', 'poder' –estos son algunos de los términos recurrentes a través de estos ensayos; nos mostramos muy curiosos ahora sobre aquellos puntos donde el arte parece ser menos privado que social» (1983: iii)².

¹ "Originally the Canon meant the choice of books in our teaching institutions, and despite the recent politics of multiculturalism, the Canon's true question remains: What shall the individual who still desires to read attempt to read, this late in history?" (Bloom, 1995: 15). Todas las traducciones son mías.

² "Interest in canons is surely part of a larger inquiry into the institutions of literary studies and artistic production. 'Politics', 'economics', 'social', 'authority', 'power' –these are some

En 1999, Nel van Djik analiza la formación del canon desde diferentes perspectivas: el nacionalismo, la literatura y el punto de vista institucional. Para él:

la lista de obras que constituye la herencia literaria de la sociedad occidental ya no es prescrita por la iglesia o el Estado, sino que por instituciones autorizadas tales como la crítica y la educación literarias. Asimismo, el desarrollo científico de las últimas décadas se ha traducido en una convicción generalizada de que el valor literario no es un valor intrínseco sino que una cualidad atribuida. Esta cualidad resulta del consenso que existe entre los miembros de una institución literaria en un determinado momento (121)³.

Según van Djik, los investigadores que trabajan sobre este tema se dividen en dos grupos: aquellos que lo abordan desde un punto de vista ideológico-hermenéutico y aquellos que optan por un enfoque sociológico.

Un año más tarde, Anderson y Zanetti (2000) determinan que el debate en torno al problema del canon se ha efectuado según dos polos opuestos. Por un lado, aquellos situados a la derecha del espectro político defienden un canon tradicional. Por otro lado, aquellos que poseen ideas de izquierda declaran que el canon es un artefacto obsoleto. Entre ambos, se sitúan una serie de posturas que abren la noción de canon con el propósito de incluir a las minorías o que promueven la existencia de múltiples cánones.

En el año 2003, Jeffrey Insko retoma el debate sobre este concepto, haciendo resurgir la tensión entre un canon imaginario y un canon pedagógico. Al año siguiente, Frank Kermode, en su libro *Placer y cambio: la estética del canon* (2004) deja de lado los aspectos ideológicos de la constitución del canon para introducir tres nuevas propiedades en dicho proceso: placer, cambio y azar.

Mi intención, al mencionar estos diversos estudios sobre el concepto del canon, no es continuar este debate, sino mostrar que la constitución de un corpus canónico es un proceso altamente complejo que no deja indiferente a nadie. Asimismo, me propongo introducir una metodología alternativa para identificar aquellas obras que son más frecuentemente analizadas al interior de una literatura nacional⁴. Es decir, en vez de presentar un canon literario, mi objeto será trazar una cartografía literaria. Específicamente, mi análisis se centra en las letras de Hispanoamérica.

ENFOQUES TEÓRICO Y METODOLÓGICO

Con el propósito de desarrollar este método experimental, baso mi investigación en la articulación de varias nociones esenciales. Desde el punto de vista

of the terms that recur throughout these essays; we are most curious now about those points where art seems less private than social" (von Halberg, 1983: iii).

³ "The list of works that count as our western society's literary inheritance is no longer prescribed by the church and the state, but by authorized institutions such as literary criticism and literary education. In addition, scientific developments in the past decades have resulted in the widespread conviction that literary value is not an intrinsic but an attributed quality. This quality results from the consensus that exists between the members of a literary institution at a certain moment" (van Djik, 1999: 121).

⁴ En su libro *Dialogues with/and Great Books* (2012), David Fishelov utiliza indicadores bibliométricos y Webométricos no con el propósito de proponer un método alternativo a la constitución del canon, sino que para confirmar la importancia de las listas canónicas ya existentes.

teórico, fundo mi trabajo, por un lado, en el concepto de campo literario establecido por Pierre Bourdieu (1992), y, por otro lado, en la cienciometría (Price, 1963; Garfield, 1980; Leydesdorff, 1998). Metodológicamente, dos enfoques resultan medulares: en primer lugar, el análisis de campo del conocimiento (Albrechtsen, 1997; Hjørland, 2001; Hjørland y Albrechtsen, 1995) y, en segundo lugar, las teorías de las citas (Kaplan, 1965; Moravcsik y Murugesan, 1975; Gilbert, 1977; Small, 1978, 1998; Leydesdorff, 1998; Enger, 2009).

Si bien varios cientometristas se muestran reacios a la utilización de técnicas cuantitativas para analizar el comportamiento de las humanidades debido a las diferencias en cuanto a las formas de citar a través de las disciplinas (Cole, 1983; Cozzens, 1985; Larivière *et al.*, 2006; Nederhof *et al.*, 1989), recientemente han aparecido numerosos análisis bibliométricos relativos a las humanidades, en general, y a los estudios literarios, en particular (Ardanuy *et al.*, 2009; Hammarfelt, 2011; Hérubel y Goedeken, 2000; Osca-Lluch y Haba, 2005). Estas últimas publicaciones indican que, a pesar de las diferencias antes mencionadas, los estudios cientométricos constituyen un enfoque pertinente para aumentar nuestra comprensión del campo literario.

Por ello, en los últimos años, he llevado a cabo diversos estudios cuantitativos relativos al área de las humanidades que me han permitido, entre otras cosas, crear un atlas de los recursos bibliográficos digitales, así como una cartografía de los estudios sobre la literatura hispanoamericana (Ferrer, 2012).

En el presente estudio, la base de datos utilizada es la Modern Language International Bibliography, que de ahora en adelante llamaré MLAIB. Esta bibliografía electrónica, mundialmente reconocida como la más importante en el área de los estudios literarios, contiene más de 2.200.000 referencias e incluye aproximadamente 4.400 revistas académicas. Además de los artículos, la base MLAIB contiene las referencias de libros, capítulos de libros y tesis⁵. En cuanto a su cobertura cronológica, ésta se extiende de 1896 al presente.

A través de las técnicas de *data mining* (Han *et al.*, 2012) y de palabras clave (Callon *et al.*, 1993), obtengo, en primer lugar, una muestra de la bibliografía crítica sobre la literatura hispanoamericana. La fecha límite utilizada es el año 2010.

CARTOGRAFÍA LITERARIA

Los resultados de esta exploración arrojaron una muestra de más de 75.000 documentos. En la figura 1, podemos observar dicha bibliografía clasificada en las diecinueve literaturas nacionales que constituyen Hispanoamérica. Adicionalmente a las referencias compiladas en el gráfico, hay un subconjunto de más de 11.000 referencias sobre las letras hispanoamericanas sin mencionar país alguno. Puesto que en este estudio me concentro en las literaturas nacionales, he omitido dichas referencias. Como podemos apreciar, la importancia relativa de los diferentes países es muy variada. Argentina y México representan cada uno 22% de la muestra y sus literaturas se destacan claramente como las más estudiadas según la base MLAIB. Luego, Cuba, Chile, Perú y Colombia muestran una bibliografía relativamente importante con muestras que fluctúan entre 12%

⁵ En este estudio, no considero las tesis, puesto que la MLAIB incluye sólo las tesis de las universidades norteamericanas. Ver Fitz-Enz (2008).

Bibliografía por literatura nacional MLAIB 1896-2010

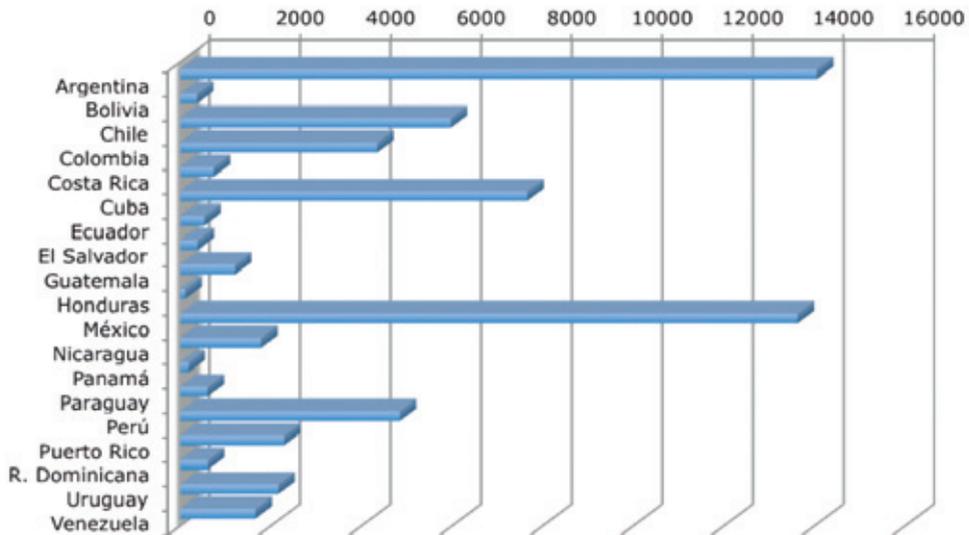


Fig. 1. Bibliografía crítica de la literatura hispanoamericana por literatura nacional.

y 7% del total. Las bibliografías sobre las literaturas de Puerto Rico, Uruguay, Nicaragua, Venezuela y Guatemala acumulan muestras pequeñas que se sitúan entre 2.300 y 1.200 referencias. Por último, Costa Rica, República Dominicana, Paraguay, Ecuador, El Salvador, Bolivia, Panamá y Honduras reflejan una situación bastante embrionaria con bibliografías que oscilan entre 800 y 100 referencias por país.

ESCRITORES Y OBRAS

Ahora bien, con el propósito de identificar las obras y los autores más estudiados de Hispanoamérica, interrogué nuevamente la base MLAIB. En la figura 2, represento, por país, el número de autores y el número de obras que reúnen 10 o más referencias dentro de la base de datos. Llegamos así a identificar 717 autores y 451 obras. Desde luego, Argentina y México siguen ubicándose en primera fila según ambos indicadores.

En el cuadro 1, presento una lista con los 20 escritores hispanoamericanos que registran el mayor número individual de referencias en la base MLAIB. Nuevamente, observamos que Argentina y México se destacan cada uno con 4 autores, seguidos muy de cerca por Cuba y Perú cada uno con 3 autores. Cabe mencionar que los seis escritores hispanoamericanos galardonados con el Premio Nobel de literatura se encuentran presentes. Sin embargo, el líder absoluto de este indicador es Jorge Luis Borges, quien nunca fuera reconocido por la Academia sueca. Asimismo, apreciamos la coexistencia de autores de distintas épocas, de Juana Inés de la Cruz a Mario Vargas Llosa. Obviamente, aquellos

**Número de autores y de obras con 10 o más referencias
MLAIB 1896-2010**

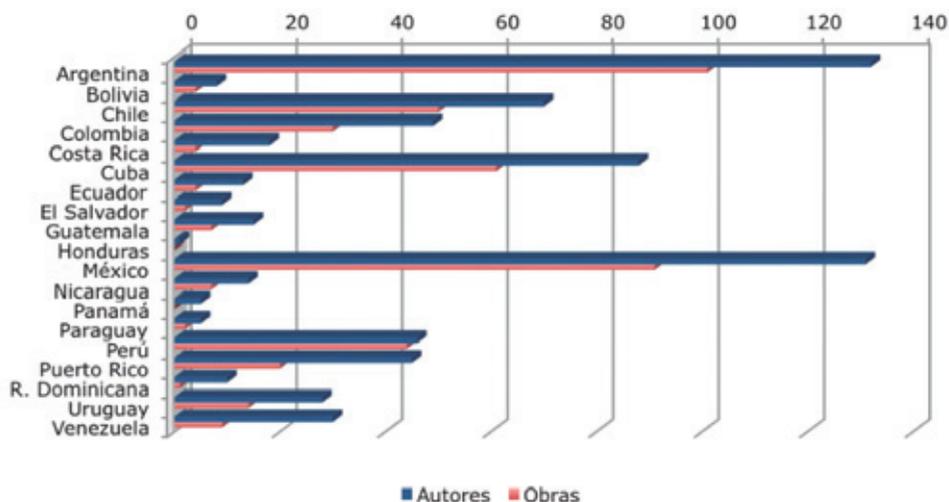


Fig. 2. Número de autores y de obras por literatura nacional.

Cuadro 1. Escritores hispanoamericanos con 100 o más referencias

País	Autor
Argentina	Borges, Jorge Luis (1899-1986)
Colombia	García Márquez, Gabriel (1928-)
Argentina	Cortázar, Julio (1914-1984)
México	Fuentes, Carlos (1928-2012)
Nicaragua	Darío, Rubén (1867-1916)
Chile	Neruda, Pablo (1904-1973)
México	Paz, Octavio (1914-1998)
Cuba	Martí, José (1853-1895)
Cuba	Carpentier, Alejo (1904-1980)
Perú	Vargas Llosa, Mario (1936-)
México	Juana Inés de la Cruz (1648-1695)
México	Rulfo, Juan (1918-1986)
Perú	Vallejo, César (1892-1938)
Guatemala	Asturias, Miguel Ángel (1899-1974)
Argentina	Puig, Manuel (1932-1990)
Perú	Arguedas, José María (1911-1969)
Chile	Mistral, Gabriela (1889-1957)
Argentina	Sarmiento, Domingo Faustino (1811-1888)
Uruguay	Onetti, Juan Carlos (1909-1994)
Cuba	Lezama Lima, José (1912-1976)

autores más recientes tienen una probabilidad relativamente baja de encontrarse en esta lista, puesto que no han alcanzado aún a acumular un número significativo de referencias. De hecho, sólo dos de los escritores de esta lista se encuentran aún en vida: Gabriel García Márquez y Mario Vargas Llosa. Por último, entre estos 20 escritores, observamos la presencia de sólo dos mujeres: Juana Inés de la Cruz, mexicana, y Gabriela Mistral, chilena.

En el cuadro 2, encontramos las obras hispanoamericanas que poseen 100 o más referencias en la base MLAIB. *Cien años de soledad*, del escritor colombiano Gabriel García Márquez, se destaca claramente como la obra preferida por los académicos. Como podemos constatar, la lista incluye obras desde el siglo XVII, tal como *Comentarios reales de los incas* de Garcilaso de la Vega, hasta los años ochenta. Al igual que en el caso de los autores, debido al efecto acumulativo de la base bibliográfica, las obras más recientes se encuentran difícilmente presentes en esta lista. En este caso, hay tres obras de escritoras: Juana Inés de la Cruz, mexicana, Isabel Allende, chilena, y Rigoberta Menchú, guatemalteca galardonada en 1992 con el premio Nobel de la Paz.

Cuadro 2. Obras hispanoamericanas con 100 o más referencias

País	Autor	Obra	Ref.
Colombia	García Márquez, Gabriel	<i>Cien años de soledad</i> (1967)	412
México	Rulfo, Juan	<i>Pedro Páramo</i> (1955)	270
Argentina	Cortázar, Julio	<i>Rayuela</i> (1963)	184
Argentina	Puig, Manuel	<i>El beso de la mujer araña</i> (1976)	164
Paraguay	Roa Bastos, Augusto	<i>Yo el Supremo</i> (1974)	152
Argentina	Sarmiento, Domingo Faustino	<i>Facundo</i> (1845)	150
Guatemala	Menchú, Rigoberta	<i>Me llamo Rigoberta Menchú</i> (1985)	149
Perú	Garcilaso de la Vega, el Inca	<i>Comentarios reales de los incas</i> (1609)	147
Colombia	García Márquez, Gabriel	<i>Crónica de una muerte anunciada</i> (1981)	137
Chile	Allende, Isabel	<i>La casa de los espíritus</i> (1982)	132
Perú	Poma de Ayala, Felipe Huamán	<i>El primer nueva corónica y buen gobierno</i> (1615-16)	124
Cuba	Carpentier, Alejo	<i>Los pasos perdidos</i> (1953)	118
Cuba	Cabrera Infante, Guillermo	<i>Tres tristes tigres</i> (1967)	116
México	Juana Inés de la Cruz	<i>Primero sueño</i> (1680-81)	116
México	Fuentes, Carlos	<i>Terra Nostra</i> (1975)	115
Cuba	Lezama Lima, José	<i>Paradiso</i> (1966)	113
Cuba	Carpentier, Alejo	<i>El reino de este mundo</i> (1949)	109
México	Fuentes, Carlos	<i>La muerte de Artemio Cruz</i> (1962)	103
Argentina	Hernández, José	<i>Martín Fierro</i> (1872, 1879)	102

El cuadro 3 corresponde a los escritores que acumulan el mayor número de obras con 10 o más referencias en la base MLAIB. A nivel de país, México presenta 6 autores, seguido por Chile con 5 autores. Individualmente, el líder es nuevamente Jorge Luis Borges. Al igual que en los otros casos, el período cubierto se extiende del siglo XVII, con Juana Inés de la Cruz, hasta el presente. En este caso, el número de escritoras alcanza a 6: las chilenas Isabel Allende y Diamela Eltit, las mexicanas Juana Inés de la Cruz, Elena Poniatowska y Carmen Boullosa y la portorriqueña Rosario Ferré.

Cuadro 3. Escritores con el mayor número de obras con 10 o más referencias

<i>País</i>	<i>Autor</i>	<i>Nº de obras</i>
Argentina	Borges, Jorge Luis	24
Perú	Vargas Llosa, Mario	18
México	Fuentes, Carlos	16
Colombia	García Márquez, Gabriel	15
Argentina	Cortázar, Julio	14
Cuba	Carpentier, Alejo	13
Chile	Eltit, Diamela	8
Chile	Neruda, Pablo	8
Chile	Allende, Isabel	7
Cuba	Martí, José	7
Uruguay	Onetti, Juan Carlos	7
Argentina	Puig, Manuel	7
México	Juana Inés de la Cruz	6
Cuba	Arenas, Reinaldo	5
Perú	Arguedas, José María	5
Chile	Bolaño, Roberto	5
México	Boullosa, Carmen	5
Chile	Donoso, José	5
Puerto Rico	Ferré, Rosario	5
México	Paz, Octavio	5
México	Poniatowska, Elena	5
México	Rulfo, Juan	5
Cuba	Sarduy, Severo	5
Perú	Vallejo, César	5

REFLEXIONES FINALES

Al inicio de este artículo, evoqué diversos aspectos que, según los académicos, dificultan el proceso de identificación de las obras literarias canónicas. Luego, introduje una metodología alternativa que se basa, esencialmente, en la minería

de datos. De esta forma, logré obtener un corpus de obras que revela el interés que, desde hace más de cien años, la crítica académica ha manifestado por la literatura hispanoamericana. A mi modo de ver, esta nueva metodología permite sobrepasar las diferentes tensiones relativas al proceso de constitución del canon y, al mismo tiempo, responde tanto a lo planteado por Nel van Dijk, cuanto a la interrogante de Harold Bloom.

Por un lado, según van Dijk, el canon es prescrito por "instituciones autorizadas tales como la crítica y la educación literarias" (1999: 121)⁶. En este sentido, el método aquí propuesto es justamente una forma de considerar una parte muy significativa de la actividad crítica académica. Al utilizar la base MLAIB y explotarla con métodos cuantitativos, la bibliografía obtenida no está sujeta a criterios de selección subjetivos, sino que se basa en la ley de los grandes números. En efecto, en el caso de Hispanoamérica, los resultados aquí compilados son el reflejo de más de 75.000 referencias obtenidas con técnicas cuantitativas reproducibles y verificables.

Por otro lado, la cartografía presentada responde a la interrogante de Harold Bloom. Es decir, adecuándola a la zona geográfica en cuestión, "¿qué debería tratar de leer un individuo que desea leer [la literatura de Hispanoamérica] a estas alturas de la historia?" (1995: 15)⁷. En este sentido, me gustaría hacer notar que los veintiocho escritores hispanoamericanos que Bloom incluye en su libro sobre el canon occidental están presentes en esta nueva cartografía.

Sin embargo, considero que la bibliografía obtenida es claramente superior a la lista canónica de Bloom, puesto que, en primer lugar, estos resultados pueden ser clasificados por número de citas y ello según diferentes parámetros: geográfico, cronológico y lingüístico, entre otros. En segundo lugar, esta cartografía es dinámica, ya que es fácilmente actualizable mediante la interrogación periódica de la base MLAIB. Por último, el corpus obtenido es claramente más rico y diverso que la lista canónica propuesta por Bloom que, en el caso de Hispanoamérica, se compone de sólo treinta y seis obras y diecisiete autores pertenecientes a ocho países. Dicha lista, que data de 1995, es claramente más restringida que las cuatrocientos cincuenta y una obras, los setecientos diecisiete autores y los diecinueve países que constituyen la cartografía propuesta. Cabe mencionar que la célebre lista de Bloom se concentra en el siglo XX y sólo considera una obra de fines del siglo XIX. Por lo demás, dicho canon no incluye a ninguna escritora, puesto que Bloom incluye a la incontestablemente sobresaliente Juana Inés de la Cruz en el canon literario de España de lo que él denomina "La era aristocrática" (1995: 503).

A mi parecer, este estudio es una demostración más de la pertinencia y la importancia de introducir métodos cuantitativos en los estudios literarios como una manera de incrementar y de profundizar el conocimiento relativo a nuestra disciplina. Por ello, considero fundamental el desarrollo de las humanidades digitales.

⁶ "by authorized institutions such as literary criticism and literary education" (van Dijk, 1999: 121).

⁷ "What shall the individual who still desires to read [Spanish American literature] attempt to read, this late in history?" (Bloom, 1995: 15).

AGRADECIMIENTOS

Esta investigación contó con el apoyo del Consejo de Investigación en Ciencias Sociales y Humanidades de Canadá (Social Sciences and Humanities Research Council of Canada).

BIBLIOGRAFÍA

- Albrechtsen, Hanne, "Knowledge organization in the humanities", *Knowledge Organization*, 24, 2 (1997), pp. 61-63.
- Anderson, Earl R. y Zanetti, Gianfrancesco, "Comparative Semantic Approaches to the Idea of a Literary Canon", *The Journal of Aesthetics and Art Criticism*, 58, 4 (2000), pp. 341-360.
- Ardanuy, Jordi; Urbano, Cristóbal y Quintana, Lluís, "A Citation Analysis of Catalan Literary Studies (1974-2003): Towards a Bibliometrics of Humanities Studies in Minority Languages", *Scientometrics*, 8, 12 (2009), pp. 347-366.
- Bloom, Harold, *The Western Canon: the Books and School of the Ages*, New York, Riverhead, 1995.
- Bourdieu, Pierre, *Les règles de l'art. Genèse et structure du champ littéraire*, París, Seuil, 1992.
- Callon, Michel; Courtial, Jean-Pierre y Penan, Hervé, *La Scientométrie, Que Sais-Je?*, París, Presses Universitaires de France, 1993.
- Cole, Stephen, "The Hierarchy of the Sciences", *American Journal of Sociology*, 89, 1 (1983), pp. 111-139.
- Cozzens, Susan E., "Using the Archive – Price, Derek Theory of Differences among the Sciences", *Scientometrics*, 7, 3-6 (1985), pp. 431-441.
- Enger, K. Brock, "Using citation analysis to develop core book collections in academic libraries", *Library & Information Science Research*, 31, 2 (2009), pp. 107-112.
- Ferrer, Carolina, "Los estudios literarios hispanoamericanos y las bases de datos digitales en la era de la globalización", en *América Latina, globalidad e integración*, Madrid, Ediciones del Orto, 2012, pp. 635-646.
- Fishelov, David, *Dialogue with/and Great Books. The Dynamics of Canon Formation*, Brighton, Sussex Academic Press, 2010.
- Fitz-Enz, Daniel, *MLA International Bibliography Database Guide*, CSA Illumina, 2008, <www.csa.com> [26/06/2013].
- Garfield, Eugene, "Is Information-Retrieval in the Arts and Humanities Inherently Different from that in Science – Effect that ISIS-Citation-Index-for-the-Arts-and-Humanities is Expected to Have on Future Scholarship", *Library Quarterly*, 50, 1 (1980), pp. 40-57.
- Gilbert, G. Nigel, "Referencing as Persuasion", *Social Studies of Science* 7, 1 (1977), pp. 113-122.
- Guillory, John, "Canon", en Joseph Childers y Gary Hentzi (eds.), *Columbia Dictionary of Modern Literary and Cultural Criticism*, New York, Columbia University Press, 1995, pp. 233-249.
- Hammarfelt, Bjorn, "Interdisciplinarity and the Intellectual Base of Literature Studies: Citation Analysis of Highly Cited Monographs", *Scientometrics*, 86, 3 (2011), pp. 705-725.
- Han, Jiawei; Kamber, Micheline y Pei, Jian, *Data Mining. Concepts and Techniques*, Waltham, Morgan Kaufmann, 2012.
- Hérubel, Jean-Pierre V.M. y Goedeken, Edward A., "Metadisciplinarity, Belles Lettres, and André Malraux: A bibliometric exploration of knowledge formation", *Serials Librarian*, 37, 4 (2000), pp. 51-68.
- Hjorland, Birger y Albrechtsen, Hanne, "Toward a New Horizon in Information-Science – Domain-Analysis", *Journal of the American Society for Information Science*, 46, 6 (1995), pp. 400-425.

- Hjorland, Birger, "Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance", *Journal of the American Society for Information Science and Technology*, 52, 9 (2001), pp. 774-778.
- Insko, Jeffrey, "Generational Canons", *Pedagogy*, 3, 3 (2003), pp. 341-358.
- Kaplan, Nancy R., "The norms of citation behavior – Prolegomena to the footnote", *American Documentation* 16, 3 (1965), pp. 179-184.
- Kermode, Frank, *Pleasure and Change. The Aesthetics of Canon*, Oxford, Oxford University Press, 2004.
- Larivière, Vincent; Archambault, Éric; Gingras, Yves y Vignola-Gagné, Étienne, "The Place of Serials in Referencing Practices: Comparing Natural Sciences and Engineering with social Sciences and Humanities", *Journal of the American Society for Information Science and Technology*, 57, 8 (2006), pp. 997-1004.
- Lentricchia, Frank y McLaughlin, Thomas (eds.), *Critical Terms for Literary Studies*, Chicago, The University of Chicago Press, 1995.
- Leydesdorff, Loet, "Theories of Citation?", *Scientometrics*, 43, 1 (1998), pp. 5-25.
- Modern Language Association International Bibliography, <www.mla.org>.
- Moravcsik, Michael J. y Murugesan, Ponraj, "Some Results on Function and Quality of Citations", *Social Studies of Science* 5, 1 (1975), pp. 86-92.
- Nederhof, Anton J.; Zwan, Rolf A.; De Bruin, Renger E. y Dekker, P.J., "Assessing the Usefulness of Bibliometric Indicators for the Humanities and the Social and Behavioral-Sciences – A Comparative Study", *Scientometrics*, 15, 5-6 (1989), pp. 423-435.
- Oscá-Lluch, Julia y Haba, Julia, "Dissemination of Spanish social sciences and humanities journals", *Journal of Information Science*, 31, 3 (2005), pp. 230-237.
- Price, Derek De Solla, *Little Science, Big Science*, New York, Columbia University Press, 1963.
- Small, Henry, "Citations and consilience in science – Comments on Theories of citation?", *Scientometrics* 43, 1 (1998), pp. 143-148.
- Small, Henry, "Cited documents as concept symbols", *Social Studies of Science*, 8 (1978), pp. 327-340.
- Van Dijk, Nel, "Research into canon formation: nationalism, literature, and an institutional point of view", *Poetics Today*, 20, 1 (1999), pp. 121-132.
- Von Halberg, Robert, "Editor's introduction", *Critical Inquiry*, 10, 1 (1983), pp. iii-vi.



RESUMEN

En este artículo, propongo un método alternativo al tradicional para la constitución del canon literario. En vez de basar la determinación del canon en diferentes valores y criterios, extraigo de la base de datos *Modern Language Association International Bibliography* los autores y las obras con el mayor número de referencias. Específicamente, mi objeto de estudio es la literatura hispanoamericana. A través de la minería de datos, obtengo una muestra de más de 75.000 referencias que nos permite observar la bibliografía crítica de las diecinueve literaturas nacionales de Hispanoamérica. Este método cuantitativo arroja un corpus de 451 obras y 717 autores que son citados más de 100 veces en la base de datos analizada. Consecuentemente, esta bibliografía no es el resultado de criterios subjetivos de selección, sino que se basa en la ley de los grandes números. Asimismo, este estudio es una demostración de la pertinencia de utilizar métodos cuantitativos para analizar las bases de datos bibliográficas digitales con el propósito de profundizar nuestros conocimientos del campo literario.

Palabras clave: Canon literario, literatura hispanoamericana, minería de datos, bibliometría, métodos cuantitativos.

ABSTRACT

In this article, I propose an alternative technique to the traditional method of constitution of the literary canon. Instead of basing the determination of the canon on different values and criteria, I scrutinize the *Modern Language Association International Bibliography* database in order to determine the most cited authors and literary works. Specifically, I study Spanish American literature. Thus, through the process of data mining, I obtain a sample of over 75,000 references that allows us to observe the critical bibliography about the nineteen national literatures of the subcontinent. This quantitative technique yields a corpus of 451 titles and 717 writers that are cited more than 100 times in the database. Consequently, this bibliography is not the result of subjective selection criteria, but is based on the law of large numbers. Furthermore, this study shows that the quantitative analysis of bibliographic digital databases is an effective way to bring new light to the field of literary studies.

Keywords: Literary canon, Spanish American literature, data mining, bibliometrics, big data.

El proyecto *Epístola*: edición digital de los epistolarios de la Edad de Plata

JUANA MARÍA GONZÁLEZ GARCÍA
Universidad Internacional de la Rioja
Juanamaria.gonzalez@unir.net

1. QUÉ ES *EPÍSTOLA*

Epístola es un proyecto de investigación impulsado por la Fundación Francisco Giner de los Ríos en colaboración con la Residencia de Estudiantes, en Madrid, que está orientado a la recuperación y edición digital de la correspondencia (o epistolarios) de las personalidades más reseñables de la Edad de Plata de la cultura española (1898-1936). Este período es una etapa especialmente relevante para la cultura española y a ella pertenecen autores como Federico García Lorca, Juan Ramón Jiménez o Pedro Salinas. La edición de la correspondencia de estos y otros autores aporta datos muy necesarios para la reconstrucción de la historia cultural de este período.

El proyecto *Epístola* ha creado durante los últimos años una colección específica bajo el sello editorial de Amigos de la Residencia de Estudiantes, donde se han ido dando a conocer los trabajos del grupo de investigación, hasta ahora, fundamentalmente, en papel. Sin embargo, actualmente se está trabajando en la aplicación de las tecnologías de la información y de la comunicación (TIC) para el tratamiento documental, lingüístico y editorial de los epistolarios de la Edad de Plata y favorecer así su investigación filológica e historiográfica. Entre los objetivos planteados se ha dado prioridad a la creación y desarrollo de un soporte de difusión de este material a través de la web y al que nos referiremos, desde ahora, como "Publicador de *Epístola*".

Todo el contenido de este *Publicador* se ha generado, mediante un *Editor* con el que se realizan las tareas de edición y anotación de las cartas. Este editor permite: (i) aplicar los criterios de anotación y transcripción acordados por los miembros del grupo de investigación *Epístola* y (ii) guardar la información compartida, tal y como: personas, instituciones, movimientos, obras, revistas, lugares y eventos que aparecen en los textos, que se consideren reseñables para el estudio de la Edad de Plata. La mayor potencialidad de la herramienta es que permitirá visualizar las relaciones entre todos estos elementos en función de las necesidades del lector o investigador a modo de gráfico interactivo. El sistema

JUANA MARÍA GONZÁLEZ GARCÍA

«El proyecto *Epístola*: edición digital de los epistolarios de la Edad de Plata»
Humanidades Digitales: desafíos, logros y perspectivas de futuro. Janus, Anexo 1 (2014), pp. 197-208

tiene características semánticas pues los elementos quedan identificados por sus metadatos y relaciones.

Tanto el "Publicador" como el "Editor de Epístola" se sirven de los resultados de otros proyectos tecnológicos desarrollados por la Fundación Francisco Giner de los Ríos y la Residencia de Estudiantes, como por ejemplo los portales "Edad de Plata" <www.edaddeplata.org> y "Revistas de la Edad de Plata" <http://www.edaddeplata.org/revistas_edaddeplata/index.html>.

2. CÓMO FUNCIONA

La edición filológica de un epistolario consiste en dos tareas fundamentales: (i) transcribir y normalizar el documento según el uso actual, respetando al máximo las peculiaridades tipográficas o de estilo de su autor, y (ii) añadir las notas de contexto que sean necesarias para aclarar los datos que se contienen en el mismo. El *Editor de Epístola* es un recurso informático que facilita ambos trabajos.

Esta herramienta, desarrollada como un complemento para Microsoft Word, añade una plantilla que contiene los estilos que se utilizarán en la transcripción de las cartas. Además incluye macros para adaptaciones específicas de la interfaz de usuario y para intermediar con un conjunto de repertorios externos de referencias, es decir, con unas bases de datos con información referida a las categorías antes mencionadas: personas, instituciones, movimientos, obras, revistas, lugares y eventos, que facilitarán el proceso de anotación de las mismas.

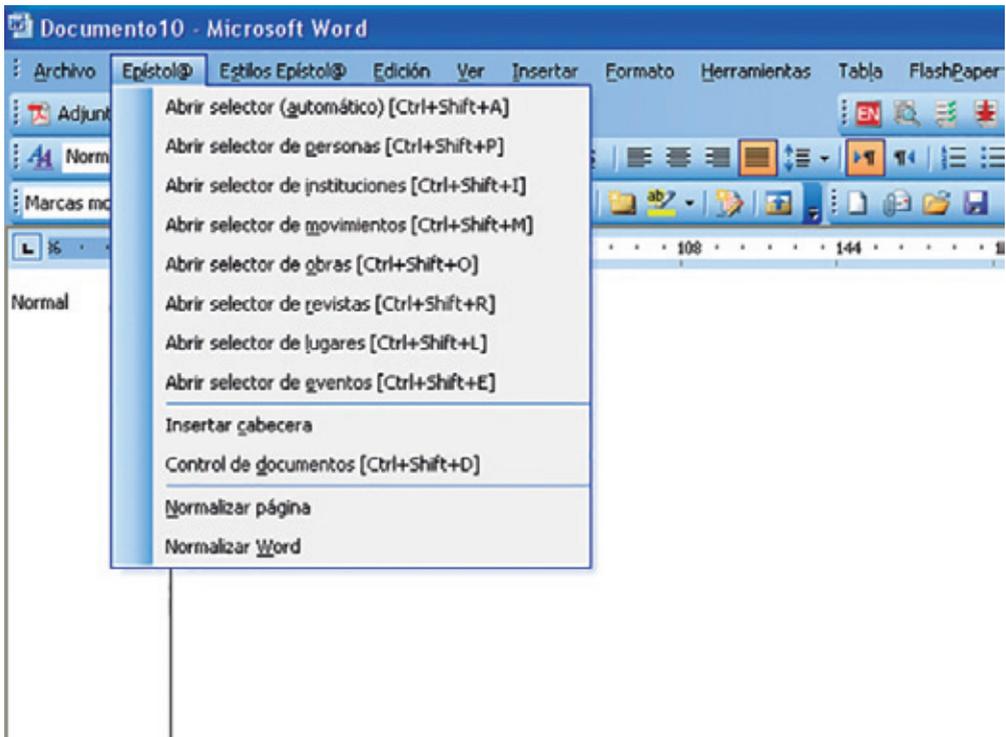
Todo investigador perteneciente al proyecto que disponga de la plantilla de *Epístola* tendrá acceso mediante protocolo web (http) a todos los documentos incorporados a la red, así como a las notas incluidas por los especialistas. Esto le dará la posibilidad, de consultar tanto los datos codificados en el texto como las relaciones que dichas recopilaciones hayan generado entre sí, así como de corregirlas o ampliarlas. De esta manera, los miembros del proyecto realizarán sus trabajos de edición *online*, lo que facilitará que la información codificada o anotada quede incorporada a la herramienta directamente, y que los investigadores puedan realizar su trabajo e intercambiar impresiones entre sí desde su lugar habitual de trabajo.

El Editor de Epístola añade dos nuevas pestañas a la interfaz de Word: *Epístola@* y *Estilos de Epístola@*, donde se encuentran se las utilidades propias del programa.

En el **menú *Epístola@*** quedan agrupadas cinco funciones principales:

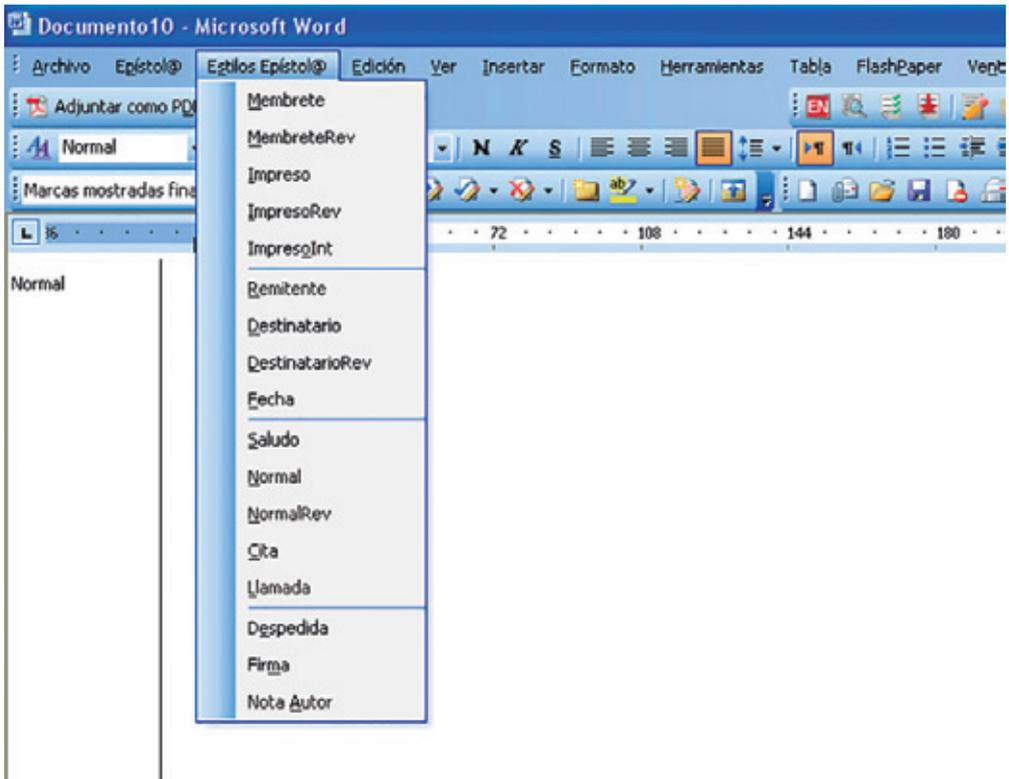
- **Selectores:** conectan el texto de la carta con las bases de datos que facilitarán las tareas de anotación y contextualización de los elementos seleccionados: personas, instituciones, movimientos, obras, revistas, lugares y eventos.
- **Control de documentos:** donde se agrupan las funciones relacionadas con la organización y clasificación de los textos, y con la organización de los documentos en el repertorio común de la web, en este caso llamado *Control de documentos remotos*. Para permitir el control de documentos será necesario introducir una serie de datos en la cabecera de cada documento (Función *Insertar cabecera*).

- **Publicación:** la función *Subir documento* se ocupa de enviar los documentos editados y terminados a la web, en la que quedarán almacenados para su publicación y consulta.
- **Edición formal:** las funciones *Normalizar Word* y *Normalizar Página* son las encargadas de eliminar los automatismos de Word que puedan afectar al proceso de transcripción y edición del texto.

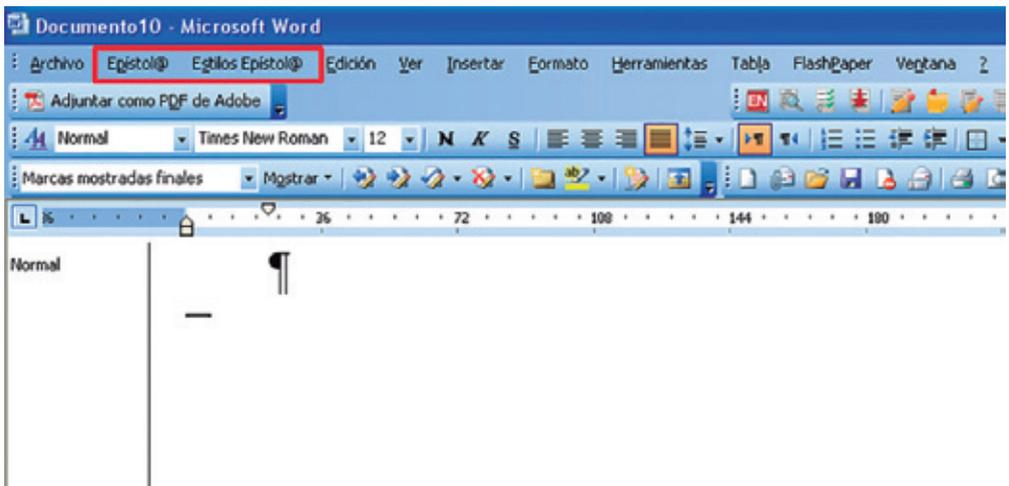


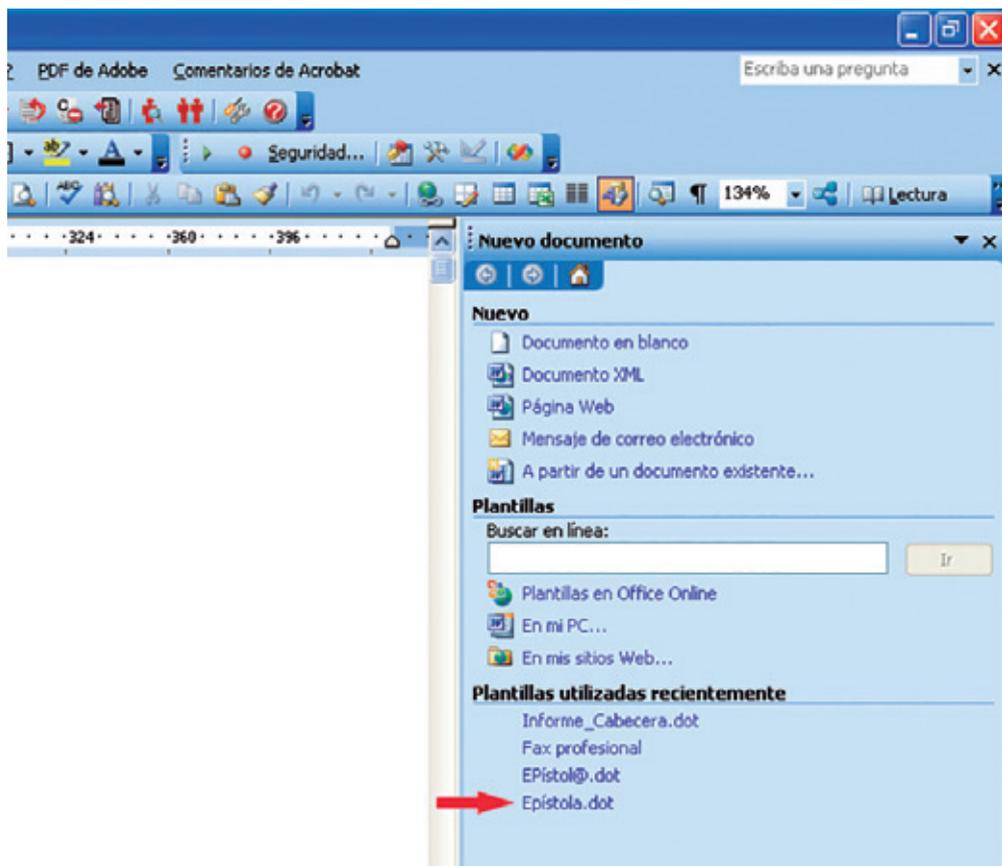
En el menú ***Estilos de Epistol@*** se encuentran sistematizados algunos de los estilos de edición del proyecto, que se emplearán en el momento de la transcripción. Éstos son, en líneas generales: Membrete, Impreso, Remitente, Destinatario, Fecha, Saludo, Normal, Cita, Despedida y Firma. Estos estilos pueden variar en función de las necesidades del investigador o características del epistolario.

Para crear un documento de *Epístola* bastará con que el usuario abra el programa de Word, seleccione "Nuevo documento" y elija la plantilla diseñada a estos efectos, en este caso denominada también *Epístola*. Ésta le dará acceso a las nuevas funciones que introduce la herramienta, con las que procederá a la transcripción y anotación del texto. Cada carta será tratada como un solo documento de Word, por lo que para trabajar con otra distinta será suficiente con la creación de un nuevo documento. Todas quedarán archivadas con un número identificador y entrarán a formar parte de una base de datos de control, a la que haremos referencia más adelante y que permitirá su clasificación y consulta sin necesidad de salir del programa.



Tras la finalización de los trabajos de edición y codificación, las cartas se exportarán mediante la función *Subir documento* al *Publicador de Epístola*, donde quedarán almacenadas y disponibles para el resto de investigadores del proyecto.





3. PARTICULARIDADES DE LA HERRAMIENTA

La edición de un documento en *Epístola* se lleva a cabo en dos áreas independientes. Por un lado, en una tabla o cabecera situada en la parte superior donde se incluyen los datos del responsable de la edición, y que aparecerá automáticamente una vez seleccionado un nuevo documento, así como los elementos de control del documento: fecha, lugar, remitente, destinatario, etc. Por otro lado, la transcripción de la carta, que se realiza en función de los estilos predefinidos a los que ya se ha aludido. El sistema automatiza de este modo la posición y características tipográficas de todo el texto, evitando posibles errores o desajustes.

3.1. Inserción y codificación de la cabecera

La cabecera del documento de *Epístola* tiene un papel fundamental en la correcta clasificación de los textos, con lo que deberá incorporarse nada más comenzar la transcripción. El investigador localizará la función "Insertar cabecera" en el menú *Epistol@* de la barra de herramientas con la que el *Editor* in-

troducirá una tabla en la que se incluyen los siguientes datos de control: *Responsable, Situación, Remitente, Destinatario, Lugar; Fecha, Comentario, Archivo, Sección, Signatura, Páginas, Tipo y Sufijo*.

The screenshot shows a Microsoft Word window titled 'NG00001 - Microsoft Word'. The table below is a representation of the data shown in the document's header area.

responsable: Juana María González=	archivo: LSC=	C
situación: =	sección: CG=	C
remitente: Pedro Salinas=	signatura: LSC/16/73A/10=	C
destinatario: León Sánchez Cuesta=	páginas: =	C
lugar: Sevilla=	tipo: Tarjeta postal manuscrita=	C
fecha: 1925.05.13=	sufijos: =	C
comentario: =		C

Una vez rellenados los campos, el investigador procederá a la codificación de algunos de los campos de la cabecera, cuestión que se explicará más adelante: en este caso, el *Remitente*, el *Destinatario* y el *Lugar*. Estos campos se corresponden con dos de las categorías incorporadas al *Editor* (persona, lugar) y facilitarán posteriormente el control de los documentos. Los elementos codificados quedarán marcados en otro color.

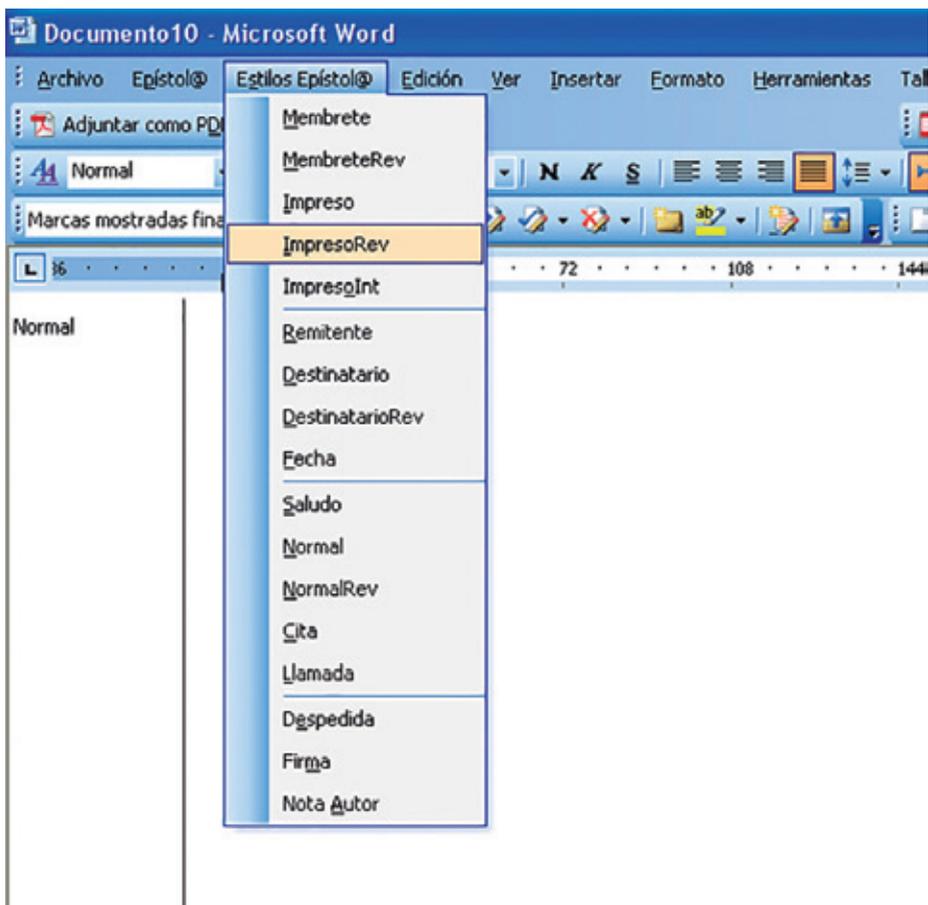
3.2. Transcripción del texto

El proyecto *Epístola* tiene establecidas, desde su creación, unas normas de edición que quedan automatizadas gracias al uso del *Editor*. Con esto se pretende conseguir una versión normalizada del documento. Estas normas de edición son dadas a conocer a los investigadores que se van incorporando al proyecto y tratan de adaptarse, en cada caso, al epistolario en el que se esté trabajando. Los estilos de *Epístola* están agrupados en la barra de herramientas según la estructura que presentan las cartas. En la parte superior, las direcciones, fechas o saludos; posteriormente, el cuerpo de la carta o contenido fundamental del documento; y finalmente, la despedida y firma.

3.3. Codificación y taxonomía de referencias

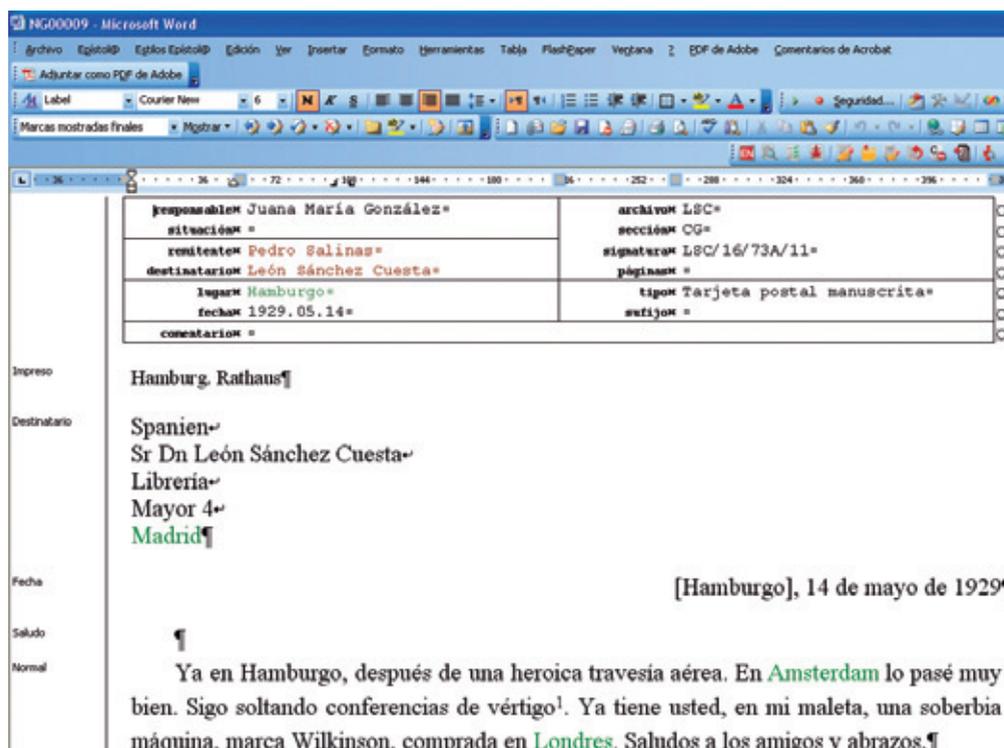
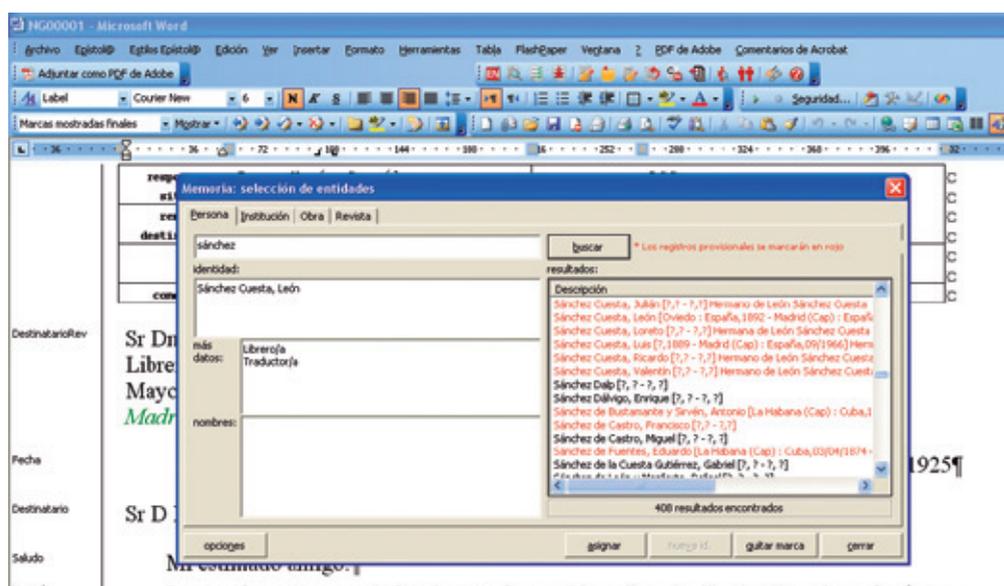
La gran novedad introducida por el *Editor* de *Epístola* reside en que reúne en una base de datos codificada toda la información dispersa en los epistolarios que el proyecto se propone editar. Este repertorio de carácter enciclopédico recibe el nombre de *Ontología de la Edad de Plata*, y está dividido en las siguientes categorías: personas, instituciones, movimientos, obras, revistas, lugares y eventos. Por ello, aparte de las tareas derivadas propiamente del trabajo de anotación, las funciones del investigador serán principalmente dos:

1. La definición y codificación de los metadatos, es decir, aquella información contenida en la cabecera de un documento de *Epístola*.
2. La codificación de los datos del texto, tales como las referencias nominales o espaciales, fechas, títulos de obras, etc., en relación con las catego-



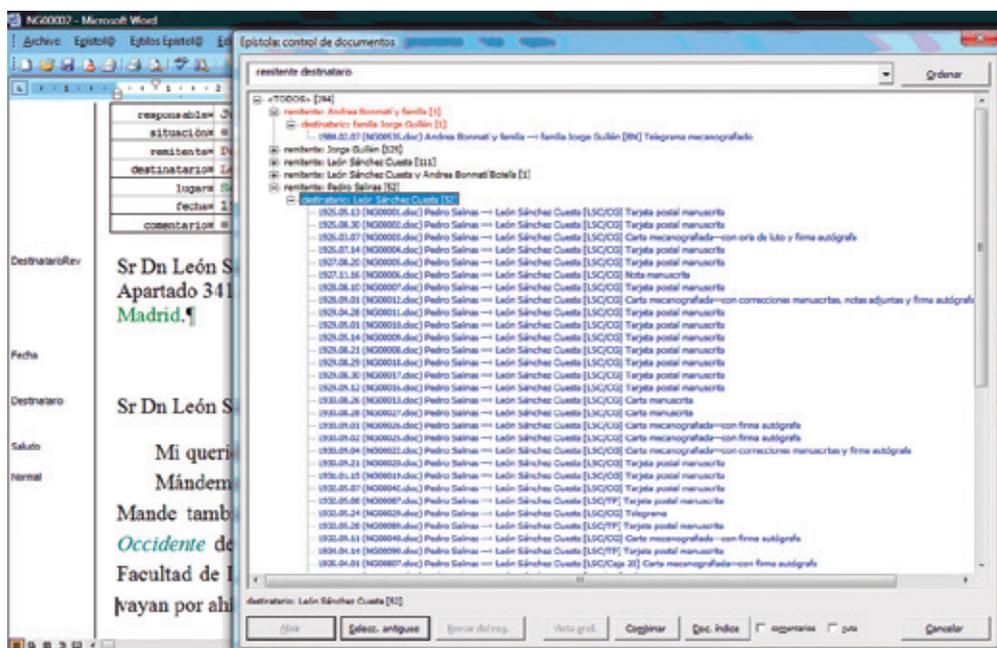
rías especificadas anteriormente: personas, instituciones, movimientos, obras, revistas, lugares y eventos, para favorecer la selección, búsqueda y acceso rápido a datos asociables que se presentan dispersos en el material original.

Por ejemplo, si en una carta apareciese el nombre de Juan Ramón Jiménez, el investigador marcaría el nombre del autor y abriría el selector de personas en el menú *Epistol@*. Si en los repertorios de referencias o base de datos referidas a personas, el registro "Juan Ramón Jiménez" estuviese dado de alta, aparecería cierta información sobre el autor (datos de nacimiento y muerte, ocupación, etc.) y el investigador se limitaría a asignarlo al elemento que se ha marcado en el texto. Si, por el contrario, el registro "Juan Ramón Jiménez" no estuviese dado de alta, el investigador tendría que crear un nuevo registro y añadir la información que la herramienta le solicitase. Posteriormente debería asignar el nuevo registro al elemento marcado en el texto. Existe la posibilidad, asimismo, de modificar o ampliar la información incorporada a un registro por otro investigador del proyecto.



3.4. Control y clasificación de los documentos

Uno de los principales problemas con los que se encuentra el editor de una correspondencia es la dificultad para mantener ordenados los documentos con un sistema de categorización eficaz. El proceso de datación de las cartas, por ejemplo, o las dificultades derivadas de la anotación del epistolario entre dos o más corresponsales, puede resultar un auténtico problema para el investigador a la hora de archivar y guardar los documentos y, así, poder consultarlos con facilidad. La codificación de los elementos de la cabecera de *Epístola* concede, sin embargo, la posibilidad de acceder a un sistema de clasificación que facilita las mismas prestaciones que una base de datos externa al programa de Word, donde los documentos pueden ser agrupados de acuerdo con los criterios que decida el investigador.

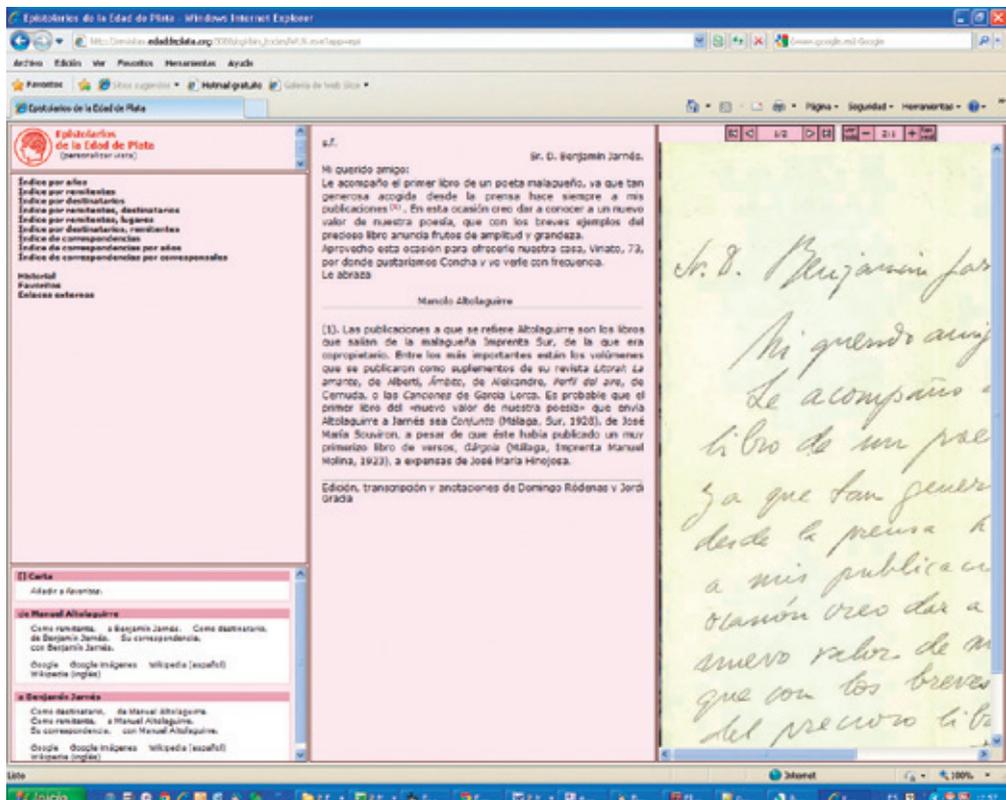


4. APARIENCIA FINAL DEL RESULTADO

Para realizar todas las tareas de consulta en la web de los documentos incluidos en *Epístola* ha sido fundamental el desarrollo de un *Publicador* dotado de un *Buscador*, que se integra en la herramienta con dos funciones diferentes para facilitar el trabajo con dichos documentos. Las tareas del *Publicador* son las siguientes:

- Acceder al *corpus* epistolar a través de diferentes índices: onomástico (remitentes, destinatarios o ambos), cronológico, etc., que se despliegan en una ventana de la herramienta a modo de árbol.

- Elegido un documento del *corpus*, dar acceso simultáneo a:
 - a) Una ventana de texto digital con la transcripción del texto, las notas del editor y las entidades marcadas en su interior, con enlaces a la información sobre las mismas.
 - b) Una ventana que permite consultar la imagen facsimilar de la pieza gracias a una herramienta de visualización.
 - c) Una ventana de *metadatos*, con la descripción completa de cada elemento del *corpus*, que permite navegar hacia cualquiera de los contenidos concretos de cada campo.
 - d) Una ventana de información complementaria, que enlaza la pieza seleccionada con la información enciclopédica recogida por el sistema o, si así lo decide el usuario, a cualquier otra entidad presente en los *metadatos* o marcada en el texto digital.



Para cotejar la transcripción con el facsímil del documento original el sistema ofrecerá una herramienta de visualización incorporada al editor y al publicador, la cual permitirá hacer *zoom* sobre la imagen con diferentes niveles de aumento, navegar por las páginas de un documento y rotar la imagen en giros de 90°.

Pieza esencial del *Publicador* es el *Buscador*, que, además de las funciones de búsqueda habituales en los metadatos y en el texto completo, explota las capacidades lingüísticas que le proporcionan el hecho de estar dotado de un completo diccionario y morfología del español que permite cosas tales como buscar todas las ocurrencias de cualquier forma verbal para un verbo determinado o buscar una sola como gerundio, verbos en tercera persona, etcétera.

5. VENTAJAS DEL PROYECTO EPÍSTOLA

Trabajar con el *Editor* y el *Publicador* de *Epístola* es muy sencillo y permitirá, en un futuro, establecer un conjunto amplio de relaciones biográficas y artísticas que posibilitará una visión original del período. Ambas herramientas poseen, potencialmente, capacidades deductivas que permitirán extraer informaciones implícitas que van más allá de la literalidad de los datos. Esta característica resultará de gran utilidad para guiar las búsquedas y relacionar informaciones en grandes conjuntos de documentación recogida de diferentes fuentes.

El proyecto *Epístola* supondrá para los investigadores una mejora considerable en el acceso a la documentación sobre la Edad de Plata, que hoy en día se encuentra dispersa y que por primera vez se está reuniendo para su edición y estudio. El proyecto favorece la puesta en común de los conocimientos individuales de los expertos en una misma materia investigadora, proporcionando una plataforma de trabajo conjunto y acumulativo que impulsa la necesaria colaboración de hispanistas de todo el mundo. De igual modo, dota a los especialistas de un sofisticado sistema de herramientas digitales integradas para el análisis de los documentos reunidos y editados.



RESUMEN

Este artículo presenta el proyecto *Epístola*, impulsado por la Fundación Francisco Giner de los Ríos en colaboración con la Residencia de Estudiantes de Madrid, que tiene como objetivo la recuperación y edición digital de los epistolarios de las personalidades más reseñables de la Edad de Plata.

El proyecto *Epístola* ha desarrollado herramientas informáticas que permiten (i) aplicar los criterios de anotación y transcripción de los textos acordados en el proyecto y (ii) guardar la información compartida entre los epistolarios anotados que se consideren reseñables para el estudio de la Edad de Plata. El sistema tiene características semánticas pues los elementos quedan identificados por sus metadatos y relaciones, lo que permitirá realizar visualizaciones de las relaciones que se establecen entre ellos y tener una nueva y original visión sobre el período.

En este artículo se explican las características de estas herramientas, sus prestaciones y su funcionamiento.

Palabras clave: Epistolario, web semántica, Edad de Plata, edición digital, correspondencia.

ABSTRACT

This article discusses the *Epístola* project. *Epístola* has been developed by the Francisco Giner de los Ríos Foundation in collaboration with the Residencia de Estudiantes in

Madrid and its objective is the recovery and digital edition of the correspondence of the key individuals from the Edad de Plata.

The Epistola project has developed tools to (i) apply a set of notation and transcription criteria agreed in the project and (ii) hold a set of records with the most relevant information associated to each letter. This creates a system with semantic characteristics, since the different elements of the letters are recorded by their metadata and inter-relations. The system will eventually enable users to visualize the relationships established between the different elements of the letters and provide new insights into the Edad the Plata.

This article explains the characteristics of these tools, their capabilities and how they operate.

Keywords: Correspondence, Semantic web, Edad de Plata, Digital edition, Letters.

Una propuesta de integración del sistema de formularios de bases de datos MYSQL con etiquetado TEI: ReMetCa, Repertorio digital de la métrica medieval castellana

ELENA GONZÁLEZ-BLANCO GARCÍA
UNED

egonzalezblanco@flog.uned.es

CLARA ISABEL MARTÍNEZ CANTÓN
UNED

MARÍA DOLORES MARTOS PÉREZ
UNED

MARÍA GIMENA DEL RÍO RIANDE
SECRIT-CONICET

proyectoremetca@gmail.com

UN REPERTORIO DIGITAL PARA LA POESÍA CASTELLANA MEDIEVAL. REMETCA

ReMetCa nace en el año 2011 como un proyecto de investigación en Humanidades de la Universidad Nacional de Educación a Distancia (2011V/PUNED/0007)¹. Se trata de una iniciativa *born digital* que se propone la creación de un repertorio métrico digital de la poesía medieval castellana, un trabajo jamás abordado hasta el momento que busca sistematizar en una base de datos consultable *online* las formas estróficas y métrico-rimáticas de todos los testimonios poéticos en nuestra lengua desde sus primeras manifestaciones en la épica a hasta la poesía cancioneril de los siglos XV y XVI, con el límite de la primera edición del *Cancionero General de Hernando del Castillo*, 1511.

El proyecto ReMetCa utiliza la tecnología más avanzada que se aplica hoy al campo de las *Digital Humanities*: los sistemas de etiquetado TEI-XML y un sistema de base de datos que permiten realizar búsquedas simultáneas a tiempo real mediante una sencilla interfaz de consulta. Este claro perfil digital que lo caracteriza no invalida que sus cimientos se basen en los estudios académicos más destaca-

¹ En la página web <www.uned.es/remetca> están detallados todos los datos del proyecto, el equipo de investigación, sus fases y el sistema de actuación desarrollado en cada momento, además de un apartado de noticias que ofrece información actualizada de congresos, eventos y proyectos de humanidades digitales vinculados con nuestro campo de trabajo.

dos sobre métrica y poesía medieval castellana². Por ello, el primer paso a la hora de sentar sus bases fue revisar los repertorios métricos impresos que comenzaron a elaborarse desde fines del siglo XIX en Europa. Cabe destacarse que una gran parte de estos trabajos encontró nueva forma en el ámbito de lo digital entre los últimos años de la centuria pasada y los primeros de este siglo.

DEL IMPRESO A LA PANTALLA: REPERTORIOS POÉTICOS 2.0

En líneas generales, los repertorios métricos surgieron en el siglo XIX a raíz de la recolección y estudio del material lírico (principalmente románico) que desde el siglo XVIII venía interesando al movimiento Romántico en Europa, pero sin lugar a dudas, fue el siglo XX el momento de la "literatura de repertorios" en papel, ya que, desde la década del 30 se perfeccionaron las diferentes técnicas de ordenamiento métrico-rimático y estrófico del material lírico. Casi a fines de la centuria, en la década del 90, asistimos a los primeros ejemplos de pasaje del repertorio impreso a la pantalla del ordenador.

Uno de los primeros repertorios que fue llevado a formato digital fue el *Die nicht lyrischen Strophenformen des Altfranzösischen* de Gotthold Naetebus (1891), repertorio métrico que reunía con un método riguroso la poesía narrativa francesa altomedieval. El *Nouveau Naetebus – Répertoire des poèmes strophiques non-lyriques en langue française d'avant 1400*, dirigido por Levente Seláf, nouveaunaetebus.elte.hu, recuperó el trabajo del filólogo alemán y lo potenció uniéndolo a otros trabajos como los de Långfors o Sinclair.

Un caso muy interesante es el del *Répertoire de la poésie hongroise ancienne*, repertorio de la poesía húngara medieval de Iván Horváth (1992), que en la década de los 90 fue pensado para ser inmediatamente digitalizado y que hoy en día muestra una última versión actualizada conforme a los avances en el ámbito digital en RPHA – *Répertoire de la poésie hongroise ancienne* (Iván Horváth et al.): <http://rpha.elte.hu/rpha/>.

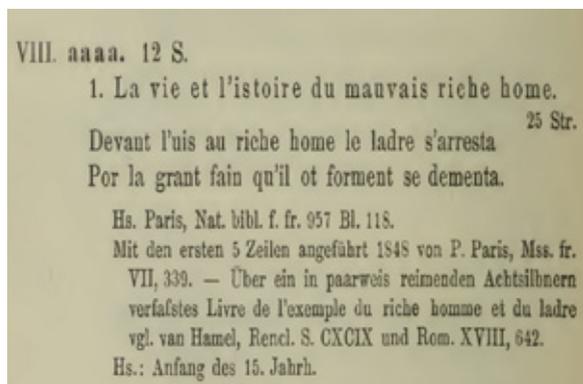
En España MedDB2, *Base de Datos de la Lírica Gallego-Portuguesa* dirigida por Mercedes Brea, <http://www.cirp.es/bdo/med/meddb.html>, recoge hoy todo el corpus profano gallego-portugués acompañado de información bibliográfica y de los esquemas métrico-rimáticos del repertorio métrico de Tavani (1967), que aquí encontró, desde nuestro punto de vista, su máxima explotación.

El repertorio métrico de Pillet y Carstens para la lírica occitana, *Bibliographie der troubadours*, de hacia 1933 (revisado por Istvan Frank hacia 1966 en su modélico *Répertoire métrique de la poésie des troubadours*), hoy puede asimismo consultarse en la BEdT – *Bibliografia Elettronica dei Trovatori*, dirigida por Stefano Asperti, www.bedt.it, donde se ha añadido a cada ficha de las piezas allí sistematizadas una gran cantidad de información filológica y bibliográfica.

Llama la atención que destacados repertorios como el de Mölk y Wolfzettel para la lírica d'oïl, *Répertoire métrique de la poésie lyrique française des origines*

² Varios integrantes de ReMetCa colaboran con el proyecto HISMETCA, *Historia de la Métrica Medieval Castellana* (financiado por el MICINN, Ref. FFI2009-09300), bajo la dirección de Fernando Gómez Redondo (UAH, Alcalá de Henares), que pronto dará a la luz los frutos de su trabajo en los volúmenes del libro *Historia de la Métrica Medieval Castellana*. Un adelanto de su contenido en: <http://www.centroestudioscervantinos.es/quienes.php?dpto=6&idbtn=1120&itm=6.1> investigación en una enciclopedia de la métrica medieval, el libro *Historia de la Métrica Medieval Castellana*.

à 1350 (1972), los repertorios de Antonelli (1984), Pagnotta (1995), Solimena (2000) o Gorni (2008) para las líricas de geografía italiana, como la siciliana y la del Dolce Stil Novo, el de repertorio de Betti (2001) para las *Cantigas de Santa María* alfonsíes³, o ya en territorio español, los pocos ejemplos de repertorios métricos que constituyen el *Repertori metric de la poesia catalana medieval* de Parramon i Blasco (1992) y el *Repertorio métrico de la poesia cancioneril del siglo XV* de Ana María Gómez Bravo (1999) no hayan sido, hasta el momento, aprovechados en una base de datos consultable *online*. Desde nuestro punto de vista esto resulta lamentable, ya que el ordenamiento que propone un repertorio métrico no está demasiado de lejos de la estructura de una simple base de datos relacional. La clasificación precisa y expresada en unos pocos datos, sumada al objetivo de reagrupar material de acuerdo a unas determinadas coordenadas, encuentra en la estructura de la base de datos y en el espacio de la pantalla de ordenador, un modo mucho más completo de aprovechamiento de las coincidencias en las búsquedas métrico-rimáticas de los repertorios en papel. Así, mientras que estas últimas limitan las coincidencias al material que vamos encontrando página a página, las realizadas en el formato digital arrojan datos que, a simple golpe de vista, pueden relacionarse entre sí. Por ejemplo, si buscásemos una estrofa amplísimamente explotada, como la cuarteta de alejandrinos monorrimos, en un repertorio en papel como el *Naetebus*, deberíamos escudriñar material a lo largo de más de una veintena de páginas. Por el contrario, esta operación en el *Nouveau Naetebus* nos arroja todas las coincidencias a partir de íncipits que se muestran en una única pantalla por la que el interesado se mueve, y de allí decide si profundizar o no su búsqueda, a partir de las diferentes fichas que se ofrecen para cada poema.



Code(x) Naetebus:	VIII.1 VIII.1.
Concordance avec les répertoires:	L'Anglois : 90 Raymond - Spanke : - Milk - Wolfzettel : -
Incipit:	Devant l'uis au riche home le ladre s'arresta
Variantes de l'incipit:	Devant l'uis au riche home le ladre s'arresta Devant l'uis au riche home le ladre s'arresta
Source:	Cambridge, Magdalene College, Coll. S. Pappas 1938 Paris, BNF, fr. 957
Titre:	(texte anonyme)
Auteur:	Cambridge, Magdalene College, Coll. S. Pappas 1938
Auteurs secondaires:	
Versification:	texte entièrement en vers poème strophique poème isostrophique
Nombre des strophes:	28 Cambridge, Magdalene College, Coll. S. Pappas 1938 Paris, BNF, fr. 957
Structure strophique:	12aaa
Irregularités métriques (numéro de la strophe - structure strophique):	8 - 12aa Paris, BNF, fr. 957 Cambridge, Magdalene College, Coll. S. Pappas 1938 10 - 12aaa Paris, BNF, fr. 957 Cambridge, Magdalene College, Coll. S. Pappas 1938 10 - 12aaa

Figs. 1 y 2. El *Naetebus* de 1891 frente al *Nouveau Naetebus* online.

³ La base de datos de las *Cantigas de Santa María*, *The Oxford Cantigas de Santa Maria Database*, <<http://csm.mml.ox.ac.uk/>>, dirigida por Stephen Parkinson, que ofrece una amplia información sobre poemas, manuscritos, milagros y distintas versiones de cada uno de los episodios, incluye asimismo información métrico-rimática para cada pieza aunque no sigue el repertorio de Betti (2001). Debe destacarse que este recurso no permite el cotejo de los distintos esquemas métrico-rimáticos de las cantigas sino que estos datos son parte de fichas que se ofrecen para cada uno de los textos del Sabio de Castilla.

En el ámbito de los estudios de la métrica medieval castellana carecemos de una herramienta digital de tipo práctico que permita realizar análisis métrico-rimáticos detallados sobre los textos, tal y como se ha hecho para las tradiciones en lengua romance que más arriba reseñábamos⁴. ReMetCa busca ser una pieza imprescindible para completar el panorama poético digital, permitiendo, gracias a las nuevas tecnologías, la realización de búsquedas potentes en varios campos de forma simultánea. Creemos que resultará, por lo tanto, de gran utilidad para estudios métricos y literarios comparativos, así como un punto de referencia para su enlace con los repertorios digitales internacionales⁵.

EL PROCESO DE CREACIÓN DE REMETCA

El paso inicial de nuestro proyecto ha sido la realización de un *webframe* o prototipo web, reflejado en la imagen que reproducimos más abajo. El objetivo de dicho prototipo fue dar respuesta a la pregunta, "¿dónde queremos llegar?" y "¿cómo queremos que sea nuestro repertorio?". Se trataba de analizar los campos de búsqueda y de resultados que considerábamos primordiales en este trabajo. La construcción del prototipo, además de constituir una tarea técnica fue también un trabajo de reflexión. Al crear el prototipo se fueron planteando una serie de problemas, muchas veces inherentes a la propia métrica medieval castellana sobre los que debimos reflexionar⁶. En primer lugar, la necesidad de realizar una distinción entre obra completa y poema, dado que una obra poética como las medievales puede ofrecer muchas variedades métricas y estróficas, como sucede, por ejemplo, con el *Siervo libre de amor*. Resulta igualmente frecuente la aparición de poemas dentro de poemas, como villancicos insertos en obras de teatro. A esto se suman otras cuestiones como la manera de representar los estribillos o versos de vuelta por la vaga delimitación de los mismos en ocasiones. Otro eje de problemas fue el derivado de la heterogeneidad de los textos literarios y del arco temporal sobre el trabajamos, en el que se incluyen desde poemas épicos como el *Cantar de Mío Cid*, que tiene tiradas muy largas en versos anisosilábicos, hasta géneros más complejos como el drama, en el que la variedad métrica se supedita a los personajes, o la poesía de cancionero, que apuesta por las innovaciones métricas.

Los criterios de búsqueda o campos buscables son, entre otros, *título*, *incipit*, *autor*, y los identificadores de otras bases de datos como *Philobiblon*⁷ y *Cancionero Virtual*⁸, aunque, como es de esperar, los campos de información métrica serán

⁴ Destacamos aquí que partimos, para la poesía cancioneril, del mencionado repertorio de Gómez Bravo (1996).

⁵ ReMetCa es parte del proyecto MEGAREP (*Megarep project*, en construcción), <<http://rpha.elte.hu/megarep/search.do>>, un megarepertorio impulsado por la Eötvös Loránd University of Sciences in the field of Medieval and Early Modern studies. Ver González-Blanco García y Sélaf (2013).

⁶ Comenzamos a sopesar problemas acerca de la metodología en el modo de codificar las rimas en las distintas tradiciones poéticas en González-Blanco García y Río Riande (2012).

⁷ Base de datos bio-bibliográfica con sede en la universidad de Berkeley sobre textos romances escritos en la Península Ibérica en la Edad Media y temprano Renacimiento, de acceso gratuito a través de Internet: <http://bancroft.berkeley.edu/philobiblon/index_es.html>.

⁸ Proyecto con sede en la Universidad de Liverpool que recoge la poesía del cancionero castellano del siglo XV basándose en las ediciones y estudios de Brian Dutton sobre la poesía cancioneril: <<http://cancionerovirtual.liv.ac.uk/main-page.htm>>.

Datación

Terminus post quem Introduzca un número de cuatro dígitos para las fechas. Si no conoce fechas exactas, ponga una horquilla aproximada o déjelo en blanco

Terminus ante quem

Clasificación de poemas [Listado de tipos de POEMA](#)

Estrofa [Listado de ESTROFAS](#)

Esquema rimático Sin comas ni espacios. Ej: **aaaa**
Minúsculas para versos de hasta 8 sílabas y mayúsculas en adelante.
Para el estribillo usamos asterisco delante del verso *a, y para los versos de vuelta el signo +a)

Esquema rimático Número de sílabas de cada verso separados por comas: Ej: 8,8,8,12

Búsqueda por rima Ej: -ida. *Esta búsqueda no es exhaustiva ya que solo se recogen rimas de las primeras estrofas de cada tipo que aparecen en los poemas

Búsqueda libre Búsquedas conjuntas entre los campos de título, incipit, autor, palabras clave, poema, estrofa

BUSCAR

Fig. 3. Captura de pantalla del prototipo del buscador de ReMetCa.

FICHA DEL POEMA Y RESULTADOS DE LA BÚSQUEDA		TEXTO
Título	Título del poema buscado	Estos versos resultaron de buscar en la REMETCA una muestra de los versos al buscar una cuarteta con criterios de la CHARTA transcribimos ciertos versos cuya edición no cuadraba con los criterios propuestos.
Otro título	Otro título por el que se conoce el poema	
Íncipit	Íncipit del poema buscado	
Autor	Autor del poema o anónimo	
Manuscrito en que se conserva el poema	Enlace a Philobiblon o en su defecto, Dutton Si no está en Philobiblon: Manuscrito en que se conserva <ul style="list-style-type: none"> • Repositorio • Idno (número, sigla) • Folios • Identificación del poema dentro del manuscrito 	Basado en la edición: Pérez Priego, <i>Teatro Medieval</i> , Madrid, Cátedra, 2009
Palabras clave		unificada según los criterios de Red Charta
género	género del listado propuesto	Enlace al texto completo en la web
tipo	tipo del listado propuesto	
tema	tema del listado propuesto	Link a la Biblioteca Virtual Cervantes o Archive... *Si el texto es breve y lo tenemos digitalizado, lo enlazaremos aquí
Datación	en formato: 1235-1345	
Tipo de poema	uno de la clasificación del listado propuesto	
Características del poema		
unisonancia	singular/unisonante	
isometrismo	heterométrico/isométrico	
isostrofismo	heteroestrófico/isoestrófico	
sonancia	consonante/asonante	

Fig. 4. Captura de pantalla del prototipo del buscador de ReMetCa.

los más relevantes para nuestro repertorio y tendrán un lugar de relevancia en los campos de búsqueda. Estos son: *esquema rimático*, *esquema métrico*, *tipo de poema*, *tipo de estrofa*, etc. A través de una búsqueda simple o compleja el repertorio arrojará, entonces, una serie de resultados en una lista que se podrá consultar además de forma individualizada para cada pieza. Esta lista de resultados está formada por diferentes fichas de los poemas. En ella encontraremos una información muy amplia: la obra completa en la que se inscribe el poema, una muestra del texto y los datos métricos concretos del poema que se ha buscado.

DECISIONES TÉCNICAS

Las características del material que debíamos sistematizar exigían una adecuada solución técnica para hacer de nuestro buscador un elemento intuitivo y útil permitiese a su usuario realizar búsquedas fácilmente, obteniendo resultados precisos. Las opciones técnicas que se nos ofrecían para realizar este tipo de repertorio eran las siguientes: una base de datos relacional o un lenguaje de marcado XML como TEI.

A grandes rasgos, una base de datos relacional que se organiza en distintas tablas entre las que se establece una relación y permite realizar búsquedas complejas presentaba una cantidad de ventajas que presenta para nuestro repertorio, circunscritas principalmente al momento en el que los investigadores o participantes del proyecto debiesen introducir datos:

- Fácil manejo: interfaz amigable para el investigador o usuario.
- Participación de colaboradores externos de forma más sencilla.
- Homogeneidad: formularios con datos finitos.

Sin embargo, esta opción presentaba asimismo una cantidad de desventajas técnicas:

- Mayor complejidad técnica inicial, que implica unos conocimientos mayores de informática no accesibles a todos los investigadores del proyecto, por lo que hace que su funcionamiento sea menos transparente.
- Dependencia de un software específico, con todo lo que ello implica desde el punto de vista de posibilidad de cambios, economía, etcétera.
- Menor flexibilidad a la hora de trabajar es formas métrico-rimáticas o estróficas complejas.

La posibilidad de utilización de un lenguaje de etiquetado o *markup language*, por el contrario, otorgaba, desde nuestro punto de vista, una mayor flexibilidad para la digitalización de un repertorio métrico. Como es sabido, un lenguaje de marcado es un tipo de lenguaje que combina texto con información extra acerca del texto por medio de las etiquetas o marcas. Las marcas también están formadas de texto, pero estas son interpretadas cuando se muestra el documento. Algunos lenguajes de marcado muy conocidos son el XML o el HTML, que se utiliza en las páginas web. El lenguaje de marcado en nuestro caso sería TEI (*Text Encoding Initiative*), <http://www.tei-c.org/index.xml>, estándar universal basado en XML para la digitalización de textos en humanidades, ciencias sociales, lingüística, etc. El módulo que más hemos desarrollado específicamente es el módulo de verso (*Verse*), que cuenta con una serie de etiquetas específicas para

la descripción de los fenómenos de tipo métrico, de las cuales hemos hecho una selección para codificar aquellos aspectos que nos interesan más de nuestros versos castellanos. Algunas de las ventajas que presentaba este sistema en el marco de nuestro repertorio métrico eran las siguientes:

- Mayor flexibilidad a la hora de codificar formas estróficas complejas.
- Disponibilidad del texto, puesto que se etiqueta sobre él.
- Acceso directo al código. Los investigadores o colaboradores que trabajen en el proyecto no lo hacen desde un formulario, sino directamente sobre el código.
- Lenguaje estándar: que hace que sea transparente para cualquiera familiarizado con TEI y que no presente complicaciones técnicas si hay cambios posteriores.
- Filosofía *open-source* y gran potencial de escalabilidad en función de las necesidades.

Como desventajas cabría añadir que para trabajar con este tipo de metalenguaje:

- Se requiere una mayor formación técnica inicial del personal investigador: necesidad de aprender el lenguaje de etiquetado.
- Necesitamos partir de una edición del texto a etiquetar fiable o, de lo contrario, el etiquetado podrá exigir casi una labor de edición.

Una vez analizadas las ventajas y desventajas de cada uno de los sistemas, nos decantamos por un sistema mixto que incluía ambas posibilidades. Nos decidimos entonces sobre la implantación del mencionado módulo *Verse* de TEI en un sistema de gestión de bases de datos relacionales y objeto-relacionales, en concreto, los sistemas de gestión de bases de datos *MySQL*. Debe destacarse que la estructura de una base de datos que a su vez incluye campos de lenguaje de marcado conjuga lo bueno de ambos sistemas, esto se traduce en un sistema de programación de bases de datos estándar y *opensource* como *MySQL*, que se combinan con campos de etiquetado XML-TEI para completar toda aquella información métrica que presenta dificultades para ser recogida mediante formularios de base de datos.

Los sistemas de *MySQL* (RDBMS) ofrecen la posibilidad de crear formularios con campos de tipo texto en los que se puede insertar marcado XML que puede luego ser recuperado mediante programas transformacionales. El almacenamiento de campos XML permite superar el tratamiento plano de los datos propio de los RDBMS, y abordar la modelización de nuestro caso de estudio, en particular, la métrica medieval castellana, con profundidad y jerarquización. Se supera la entidad relacional para situarnos en el paradigma de la orientación a objetos. La superación de lo relacional, al menos para el campo que nos ocupa, reduce los costes de tratamiento y recursos, al evitar las onerosas operaciones de combinación de tablas, tanto a la hora de la consulta como de la introducción de datos. Esto se consigue por la posibilidad, inherente al lenguaje de marcación XML, de anidar elementos. Así, un determinado elemento junto con los que a él se anidan, será traducido a un solo campo de la base de datos. Para su manipulación, por lo tanto, se recurrirá sólo a una tabla.

En efecto, los campos XML permiten un tratamiento avanzado tanto en lo que se refiere a su manipulación o a su consulta, combinando el potencial de

los lenguajes SQL, Xpath o Xquery, beneficiándonos, además, de las API's de programación, perfectamente desarrolladas, documentadas y maduras, de los sistemas relacionales de gestión de bases de datos.

Una de las opciones que hemos estudiado es el uso de la base de datos de Oracle Database 9i y 10g, que cuenta con la ventaja de poseer un campo XML-Type con un parser incluido en el mismo, gracias al componente XDB, frente al campo de mero texto de MySQL. Esta propuesta, que finalmente hemos descartado por razones de coste, la planteamos en González-Blanco García y Rodríguez (2013).

Frente al uso simple de *MySQL*, Oracle permite trabajar con esquemas XSD, de forma que el sistema puede confirmar o rechazar la validez del texto XML introducido, es decir, no sólo si el texto está bien formado, sino también si cumple las reglas establecidas en el esquema que previamente hayamos introducido. La naturaleza objeto-relacional de *Oracle* frente a la puramente relacional de *MySQL* es la que marca la diferencia en cuanto a funcionalidades, revelándose mucho más avanzada el tratamiento y manipulación de datos a través de *Oracle*.

En síntesis, las herramientas digitales que utilizamos en dicha propuesta eran:

- *Oracle Database 11g Express Edition.*
- Sistema de Gestión de Bases de Datos *MySQL*.
- *Oxygen* (editor de XML).
- *MagicDraw UML Personal Edition.*
- *MySQLWorkBench.*

Como ya hemos señalado, finalmente hemos optado por seguir trabajando con *MySQL*, ya que, a pesar de sus limitaciones, cubre perfectamente nuestras necesidades, al menos por el momento.

Reproducimos aquí un esquema del diseño de nuestra base de datos que recoge varias tablas. Las principales son la de *obra completa* y la de *poema*. En ellas se aglutina toda la información referente al autor, al manuscrito, a los datos externos, y a la información métrica de cada uno de los esquemas que aparecen en el conjunto. Así, toda obra completa tiene que tener como mínimo un poema, o esquema métrico diferente, pero también puede contener varios. Por ejemplo, en un caso como el *Libro del buen amor* que reúne distintos esquemas métricos tendríamos que rellenar varias tablas correspondientes a poema para cada uno de los distintos esquemas, pero solamente tendríamos una tabla de obra completa que recogería todos los datos de la obra en su conjunto.

Dentro de la base de datos se encuentra el campo de etiquetado TEI-XML en el que incluimos el texto que estamos analizando y reflejamos con etiquetas las particularidades métricas del mismo, como son el número de verso, el número de estrofa, la rima o el esquema métrico.

CONCLUSIONES

La importancia del proyecto ReMetCa radica fundamentalmente en su carácter innovador y multidisciplinar, ya que combina, como ya hemos señalado, la aproximación de los estudios métricos y poéticos más sobresalientes de la Filo-

Primer verso del poema :	Andemos, señora, andemos
Número de versos :	18
Número de estrofas :	5
FormulaMétrica :	8,8,8,8,8,8,8,8,8,8,8,8,8,8,8,8
FormulaRimática :	zzabbaaz*z*zcddccz*z*z
Código TEI :	<pre> <lg type="estribillo" subtype="pareado" asonancia="consonante" met="8,8" rhyme="zz"> <l n="53" rhyme="emos">Andemos, señora, andemos.</l> <l n="54" rhyme="emos">o si manda, descansemos.</l> </lg> <lg type="estrofa" subtype="redondilla" asonancia="consonante" met="8,8,8,8" rhyme="abba"> <l n="55" rhyme="ón">No me carga mi zurrón,</l> <l n="56" rhyme="ado">no he menester mi cayado,</l> <l n="57" rhyme="ado">que de Dios soy consolado,</l> <l n="58" rhyme="ón">libre de toda pasión,</l> </lg> <lg type="estrofa" real="consonante" met="8,8,8,8" rhyme="az*z*z"> <l n="59" rhyme="ón">pues que nuestra redención</l> <l n="60" rhyme="emos">con nosotros la traemos,</l> <lg type="estribillo" subtype="pareado" asonancia="consonante" met="8,8" rhyme="z*z"> <l n="61" rhyme="emos">Andemos, señora, andemos.</l> <l n="62" rhyme="emos">o si manda, descansemos.</l> </lg> <lg type="estrofa" subtype="redondilla" asonancia="consonante" met="8,8,8,8" rhyme="cddc"> <l n="63" rhyme="ero">El descanso verdadero</l> <l n="64" rhyme="oso">es nuestro hijo precioso,</l> <l n="65" rhyme="oso">Este es Dios poderoso,</l> <l n="66" rhyme="ero">Este es el manso cordero,</l> </lg> </pre>

Fig. 5. Captura de pantalla del formulario de ReMetCa.

logía Hispánica con las posibilidades más innovadores de tratamiento de datos en la actualidad. Las contribuciones que realiza, creemos, son útiles tanto para el campo de la métrica medieval como para el de las Humanidades Digitales. El proyecto fomenta además el libre y universal acceso a su base de datos y la colaboración entre investigadores de diversas latitudes, ya que se basa en la filosofía del *crowdsourcing*⁹.

En nuestra opinión, la integración del análisis filológico con la investigación en nuevas tecnologías de parte de todos los integrantes de ReMetCa abre una nueva etapa en la línea de investigación de las Humanidades Digitales dentro del área de la filología y la teoría literaria en el ámbito hispánico, proponiendo la formación tecnológica del humanista, que no sólo incorpora personal técnico a su proyecto de investigación sino que estudia, analiza y codifica junto con él el material filológico.

BIBLIOGRAFÍA

Antonelli, Roberto, *Repertorio metrico della scuola poetica siciliana*, Palermo, Centro di Studi Filologici e Linguistici Siciliani, 1984.

⁹ Una aproximación al tema de la generación de recursos humanos a través del trabajo mancomunado de personas con intereses comunes en Howe (2006).

- BEdT – *Bibliografia Elettronica dei Trovatori* (dir. Stefano Asperti), <www.bedt.it> [09/11/2013].
- Betti, Maria Pia, *Repertorio Metrico delle Cantigas de Santa Maria di Alfonso X di Castiglia*, Pisa, Pacini, 2005.
- Gorni, Guglielmo, *Repertorio metrico della canzone italiana dalle origini al Cinquecento (REMCI)*, Firenze, Cesati, 2008.
- Gómez Bravo, Ana María, *Repertorio métrico de la poesía cancioneril del siglo XV*, Alcalá de Henares, Universidad de Alcalá de Henares, 1999.
- González-Blanco García, Elena y Seláf, Levente, "Megarep: A comprehensive research tool in medieval and renaissance poetic and metrical repertoires", en L. Soriano, M. Coderch, H. Rovira, G. Sabaté y X. Espluga (eds.), *Humanitats a la xarxa: món medieval/Humanities on the web: the medieval world*, Oxford-Bern-Berlin-Bruxelles-Frankfurt am Main-New York-Wien, Peter Lang, 2013.
- González-Blanco García, Elena y Rodríguez, José Luis, "ReMetCa: a TEI based digital repertory on Medieval Spanish poetry", en F. Ciotti y A. Ciula (eds.), *The Linked TEI: Text Encoding in the Web, Book of Abstracts – electronic edition. Abstracts of the TEI Conference and Members Meeting 2013: October 2-5, Rome*, Rome edited by DIGILAB Sapienza University & TEI Consortium, 2013, pp. 178-185, <<http://digilab2.let.uniroma1.it/teiconf2013/abstracts/>> [09/11/2013].
- González-Blanco García y del Río Riande, M^a Gimena, "Uso y función del verso alejandrino en las *Cantigas de Santa María*", en *Ars metrica*, 5 (2012), <<http://ars-metrica.germ-ling.uni-bamberg.de/>> [09/11/2013].
- HISMETCA. *Historia de la métrica medieval castellana*, Fernando Gómez Redondo (dir.), Centro de Estudios Cervantinos, <<http://www.centroestudioscervantinos.es/quienes.php?dpto=6&idbtn=1120&itm=6.1>> [09/11/2013].
- Hórvath, Iván, *Répertoire de la poésie hongroise ancienne. Manuel de correction d'erreurs dans la base de données*, 2 vols., Paris, Nouvel Objet, 1992.
- Howe, J., "The rise of crowdsourcing", 2006, <<http://www.wired.com/wired/archive/14.06/crowds.html>> [09/11/2013].
- MedDB2 – *Base de datos da Lírica profana galego-portuguesa*, Mercedes Brea (dir.), <<http://www.cirp.es/bdo/med/meddb.html>> [09/11/2013].
- Mölk, Ulrich y Wolfzettel, Friedrich, *Répertoire métrique de la poésie lyrique française des origines à 1350*, Munchen, W. Fink Verlag, 1972.
- Naetebus, Gotthold, *Die nicht lyrischen Strophenformen des Altfranzösischen*, Leipzig, S. Hirzel, 1891.
- Nouveau Naetebus* – Répertoire des poèmes strophiques non-lyriques en langue française d'avant 1400 (Seláf, Levente), <nouveaunaetebus.elte.hu> [09/11/2013].
- Pagnotta, Linda, *Repertorio metrico della ballata italiana*, Milano-Napoli, Ricciardi, 1995.
- Parramon i Blasco, Jordi, *Repertori mètric de la poesia catalana medieval*, Barcelona, Curial, Abadia de Montserrat, 1992.
- Pillet, Alfred y Carstens, Henry, *Bibliographie der Troubadours*, Halle, M. Niemeyer, 1933.
- ReMetCa, *Repertorio métrico digital de la poesía medieval castellana*, Elena González-Blanco García (dir.), <www.uned.es/remetca> [09/11/2013].
- RPHA – *Répertoire de la poésie hongroise ancienne*, Iván Horváth et al. (dirs.), <<http://rpha.elte.hu/rpha/>> [09/11/2013].
- Solimena, Adriana, *Repertorio metrico dello Stil novo*, Roma, Presso la Societa, 1980.
- Solimena, Adriana, *Repertorio metrico dei poeti siculo-toscani*, Palermo, Centro di studi filologici e linguistici siciliani in Palermo, 2000.
- Tavani, Guisepppe, *Repertorio metrico della lingua galego-portoghese*, Roma, Edizioni dell'Ateneo, 1967.
- The Oxford Cantigas de Santa Maria Database*, Stephen Parkinson (dir.), <<http://csm.mml.ox.ac.uk/>> [09/11/2013].



RESUMEN

ReMetCa, *Repertorio Métrico Digital de la Poesía Medieval Castellana*, es el primer repertorio métrico *online* y de libre acceso para la poesía medieval castellana. ReMetCa es un proyecto métrico *born digital* que abarca los testimonios poéticos existentes en nuestra lengua desde sus primeras manifestaciones a finales del siglo XII (épica, cuaderna vía, etc.) hasta la riqueza y variedad de formas métricas presentes en la poesía cancioneril de los siglos XV y XVI. A pesar de que los estudios sobre métrica y poesía medieval han merecido cierto interés en los últimas décadas, carecemos de una herramienta de tipo práctico que permita realizar trabajos comparativos, tal y como lo hacen los repertorios métricos digitales de otros *corpora* poéticos como el gallego-portugués medieval (MedDB2), francés medieval (Nouveau Naetebus), occitano medieval (BedT), o mediolatino (Pedecerto), por nombrar unos pocos.

El proyecto está basado en la integración de los conocimientos métricos tradicionales de nuestra poesía (esquemas rítmicos y rimáticos, etc.) con los estándares tecnológicos que se aplican hoy al campo de las Humanidades Digitales: los sistemas de etiquetado TEI-XML combinados con un sistema de formularios MySQL que permiten realizar búsquedas simultáneas *online* mediante una sencilla interfaz de consulta. Explicaremos la evolución que nos hizo decantarnos por esta opción mixta al no resultar suficiente el sistema de tablas relacionales para cubrir la heterogeneidad formal de nuestro sistema poético. La posibilidad de añadir etiquetado TEI (y especialmente su módulo *verse*) sobre el texto nos permite, en primer lugar, poder realizar búsquedas mucho más libres a través de los esquemas métricos (sin tener que considerar como unidades cerradas los poemas o las estrofas), así como recuperar los ejemplos textuales vinculados a cada uno de los esquemas métricos o rimáticos, etc.

Palabras clave: Repertorio, métrica medieval, Humanidades Digitales, base de datos relacional, TEI-XML.

ABSTRACT

ReMetCa, *Repertorio Métrico Digital de la Poesía Medieval Castellana (Digital Repertoire on the Metrics of the Medieval Castilian Poetry)* is an online, open access metrical repertoire on Medieval Castilian poetry. ReMetCa is a born digital project that gathers poetic testimonies from the 12th century (epics, ballads, cuadernavia, etc.) up to the rich and varied poetic manifestations of the 15th-16th century Castilian *Cancioneros*. Although metrical studies on Spanish Medieval poetry have been developing fast over the past few years, researchers have not yet created a digital tool to undertake complex searches on the Castilian corpus as it has already been done with other medieval lyrical traditions in Romance languages, such as the Galician-Portuguese (MedDB), French (BedTrouveres, Nouveau Naetebus), Occitan (BedT), or the Medieval poetry in Latin (Pedecerto), just to name a few of them.

The project is based on the integration of traditional metrical and poetic studies (rhythm, rhyme patterns, etc.) with Digital Humanities, the tagging systems TEI-XML, and the creation of a database to undertake online simultaneous searches by means of a simple interface created through MySQL. We will explain here our final decision in relation to our database, a hybrid model (relational and XML) that covers our needs when working with our poems. The possibility of adding TEI code (especially its module *verse*) helped us to work on free complex searches through our schemes (without considering poems or strophes as a whole), and to retrieve the poetic texts related to those metrical or rhymatic schemes.

Keywords: Repertoire, medieval metrics, Digital Humanities, relational database, TEI-XML.

Análisis crítico de los discursos y narrativas digitales. *The Gallery of Lost Art* –Tate– como estudio de caso

M^a ISABEL HIDALGO URBANEJA
Universidad de Málaga
mbelhu@gmail.com y mbelhu@uma.es

El medio digital está permitiendo y favoreciendo un proceso de cambio en los discursos ligados a la institución museística. En paralelo a estas nuevas modalidades discursivas, surgen también las tipologías narrativas como reflejo de nuevos modos de transmitir y concebir la historia del arte y las propias obras.

El análisis que presento refleja algunas de las principales cuestiones que están centrando la temática de mi tesis doctoral¹. La elección de este caso en concreto, la exposición virtual *The Gallery of Lost Art* de la Tate², responde a su relevancia en el marco de las prácticas que se están sucediendo en la actualidad en el ámbito internacional, dentro del cual ha obtenido un sonoro reconocimiento y una gran visibilidad. Si este ejemplo ha supuesto un paso adelante en el diseño de discursos museísticos y exposiciones virtuales, su temática es igualmente novedosa al realizar un recorrido por obras del arte contemporáneo perdidas bajo diversas circunstancias, que han sido recuperadas y mostradas al público nuevamente gracias a las posibilidades de Internet.

La aplicación de los *Cultural Studies* en museología nos ha llevado a entender el museo bajo determinadas perspectivas como son el postestructuralismo y la textología (Mason, 2006), introduciendo entre su terminología conceptos propios de estas teorías como discurso y narrativa.

En primer lugar, para poder establecer un estatuto epistemológico de la museología –y por consiguiente de la museología propiamente digital–, es necesario adoptar el concepto de discurso de Foucault, que aplicado a nuestra disciplina se entendería como: una estructura de conceptos y conocimiento definidos por criterios inherentes a la propia museología que pertenecen y se implican dentro de acuerdos de la propia institución. Lyotard (1991) daría un paso más, y en su definición del término Metarrelato, con connotaciones similares a la de discurso, hace patente el cuestionamiento y relativización del peso específico de estos en nuestra sociedad al hacerse patente la existencia de otros discursos periféricos

¹ Título tesis doctoral: *Discursos museológicos y museográficos en entornos digitales*.

² <<http://galleryoflostart.com/>> [18/06/2013].

que reclaman su espacio y reconocimiento dentro de un determinado campo de conocimiento, como en este caso sería el museológico.

No se puede obviar que de igual manera, en lo que respecta al medio digital, tanto el postestructuralismo como el posmodernismo representan dos de los grandes ejes teóricos del mismo. Tomando como referencia a Lev Manovich (2005), éste señala la teoría tanto de Foucault como de Lyotard como prefiguradoras de los nuevos medios. Y precisando aún más, en lo que corresponde al análisis de la museología digital, Ross Parry (2007), uno de los principales autores que han teorizado ésta, adopta las cinco características de la teoría de Manovich como la base de partida para entender y definir el museo en su vertiente digital.

Para continuar sentando las bases de este estudio de caso, hay que tener en cuenta igualmente la perspectiva textológica que entiende el museo como un texto en el que se desarrollan una serie de discursos primordialmente bajo una estructura narrativa (Bal, 1992). Esta vertiente textológica nos insta a comprender igualmente al propio medio digital, ya que es un elemento textual, el hipertexto, el que se establece como punto de partida para la construcción, estructuración y desarrollo de las eventuales narrativas hipertextuales –o más concretamente en este caso hipermediales– en Internet.

Antes de entrar de lleno en el análisis he de precisar que lo he estructurado en cuatro niveles, respondiendo a los diferentes aspectos y componentes más relevantes de la exposición virtual: la estructura de la misma, la interfaz que ha adoptado, la posición de la institución museística como autor de ella, y la función y respuesta del usuario hacia ésta.

La primera cuestión a analizar es, como se ha referido en las líneas superiores, la estructura de esta exposición virtual tomando como punto de partida la forma de *base de datos*, bajo la que según Manovich se define en mayor medida la idiosincrasia del medio digital, pero también, como precisa Parry (2010: 13), es una forma cultural constructiva que simboliza al propio museo.

En el caso de *The Gallery of Lost Art*, nos encontramos ante una recopilación del material testimonial de las obras representadas organizado y clasificado siguiendo las diferentes circunstancias bajo las que se ha producido la pérdida de cada una de ellas, pudiendo haber sido: atacadas, robadas, destruidas, descartadas, rechazadas, irrealizadas, borradas, transitorias, efímeras, o simplemente, perdidas. Este material reunido para la exposición se presenta en formatos, como la fotografía, los recortes de prensa, las cartas o las ilustraciones en catálogos y libros, todos provenientes de diferentes archivos y bibliotecas públicos y privados.

Fruto de esta *base de datos* de obras de arte contemporáneo perdidas, es el gran discurso en torno a la materialidad artística que se desarrolla a tres bandas: por una parte, la materialidad de la obra artística vista en clave negativa que conlleva su desaparición y que provocaría mayoritariamente su marginalización y olvido dentro del discurso generalista de la historia del arte. Aunque este concepto negativo se contrapondría a la inmaterialidad concebida como recurso plástico que ha sido tan recurrente y que caracteriza al concepto de arte en contemporaneidad. Y por último, el discurso más relevante en este caso que sería el matiz digital del propio medio, algo a lo que Parry (2007: 68) define como *e-tangibilidad*, que dota a las obras desaparecidas de un nuevo estado material al entrar en contacto con el medio digital y que eventualmente puede recobrar el estatus de ciertas obras presentes en la exposición en la historia del arte.

La narración resultante de esta estructuración hipermedial de la base de datos que define Internet, en el caso de esta exposición se destaca por la utilización de ciertos conceptos que en este caso funcionan como un recurso rítmico. Y es que las obras que componen la exposición, sobre unas cuarenta, se han ido añadiendo al espacio virtual expositivo periódicamente, además de establecerse una duración anual de la exhibición (finalizada el 3 de julio de 2013), que respondería a los parámetros utilizados en las exposiciones de carácter temporal. El concepto y denominación tan recurrente en la web de *beta* –que sin ir más lejos ha llegado a ser utilizado por la propia Tate en su web institucional durante un determinado período de tiempo–, que se puede definir como la provisionalidad fruto de la continua renovación e inmediatez de Internet que conllevaría en última instancia a la obsoletización tanto de formatos como de tecnologías, y en el caso de *The Gallery of Lost Art* se podría caracterizar perfectamente su desarrollo narrativo. Se busca un desarrollo progresivo de la exposición con la adición de las obras cada cierto tiempo buscando atraer la atención del visitante de forma recurrente, además de la adicción de un contador con el tiempo restante para que la exposición desaparezca que se puede interpretar como un juego de paralelismo con las circunstancias de propias obras que contiene. De hecho, una vez finalizada la exposición podemos encontrar alojada en la url originaria una nueva web con contenido testimonial de la exposición encabezada con la frase *The Gallery of Lost Art has been erased*³.

No obstante, hay que recalcar que la duración de la exposición es fruto de una cuestión meramente presupuestaria y organizativa, como recalcan Jennifer Mundy y Jane Burton (2013), respectivamente comisaria de la exposición y responsable y creativa de *Tate Media*, los requerimientos y legislación de acuerdo con los derechos de autor y reproducción de las obras son similares a los de una exposición temporal convencional no adaptándose a los nuevos medios. Una problemática a la que se enfrentan muchos museos en la actualidad que ha contribuido en este caso concreto (junto a otros aspectos) a elevar el coste de producción a los niveles de una exposición convencional de la propia Tate.

El éxito de *The Gallery of Lost Art* consiguiendo representar una temática y obras bajo condiciones tan particulares depende en gran medida de un diseño de interfaz innovador por numerosas razones. En la línea de la identidad visual de la institución británica, siempre a la vanguardia en cuanto a diseño y calidad creativa, en este caso la Tate ha obviado la representación clásica del museo para optar por un archivo o almacén, precisamente correspondiendo la forma simbólica de la base de datos de Manovich. Se podría considerar como el último eslabón en la evolución de los modelos museísticos virtuales que buscan la reconstrucción y recreación espacial.

Igualmente novedosa ha sido la utilización de ciertos recursos ajenos a la museología digital hasta el momento, en especial la dramatización del espacio utilizada tanto el teatro o el cine tradicionalmente, que de igual manera en los últimos tiempos ha ido siendo incorporada a los videojuegos. Tanto el punto de vista cenital desde el que se distribuye el conjunto expositivo, respondiendo a la isotropía propia de los espacios virtuales, y que coloca al visitante en una posición privilegiada a la vez que irreal, a la utilización de una iluminación que marca los puntos de interés, así como la adición de un hilo musical, todos estos

³ <<http://galleryoflostart.com/>> [05/07/2013].

elementos niegan la neutralidad para ambientar y contextualizar la exposición enriqueciendo de manera notable la narración histórico-artística, han ido configurando el aspecto de lo que en palabras de Burton sería una *galería fantasma*.

Como dato curioso, hay que señalar que las referencias de los diseñadores de la interfaz provienen en gran parte de la escenografía de la película *Dogville* de Lars Von Trier –a su vez influenciada por la de los teatros “black box”– además, el equipo realizador tenía experiencia en series de tv con escenas de crímenes e investigaciones forenses, algo que se refleja claramente en el tono general de el ambiente recreado (Mundy, Burton, 2013).

La verdadera disolución de la autoría y la autoridad cognoscitiva en la red que subyace bajo el concepto postmoderno de *muerte del autor* (Barthes, 1994) es una cuestión clave sobre la que conviene analizar la relevancia real de la Tate como institución y su postura ante la historia del arte oficial con la elección de una temática tan particular que cuestiona la naturaleza material del arte en un medio igualmente inmaterial.

La Tate es uno de los grandes pilares del panorama artístico occidental –por no decir mundial– un claro ejemplo de éxito entre crítica y público dentro del circuito museístico más mainstream, siendo una de las pocas que puede permitirse presentar una exposición virtual de tales características, tanto por el alto coste de realización como por la libertad para poder afrontar una temática audaz de forma extremadamente cuidada.

En una primera aproximación deducimos que la visión de la historia del arte perdido es cuanto menos parcial a pesar de defender y remarcar la elección de las obras más importantes de la historia contemporánea. Únicamente revisando los artistas seleccionados en la exposición es manifiesto que institución podría estar detrás de su elaboración, por ejemplo, destacando una especial atención sobre ciertos nombres del arte británico, entre los que están ciertos nombres del colectivo de los *young british artists*, frente a otras nacionalidades de artistas contemporáneos vivos. De hecho, creo que si *The Gallery of Lost Art* pudiera reproducirse en un museo real, encajaría perfectamente en la programación de casi cualquiera de las sedes museísticas de la Tate.

Un hecho a recalcar es que, a pesar del peso específico de la Tate dentro del panorama museístico, es cuanto menos determinante de la verdadera consideración del medio digital entre ciertos círculos que la exposición no haya tenido demasiada acogida entre la crítica artística con mayúsculas. Esta situación según Mundy y Burton:

“It is rather a reflection of the difficulty that many in the art world have accepting the intellectual value of sites designed for mass audiences and which are associated in the minds of many with a sphere of popularising communications and transient promotion. This remains a powerful disincentive for further exploration or creative innovation in this field and needs to be addressed, we feel, if the potential of online exhibitions is to be realised and curators, scholars, and artists are to be involved”⁴.

Lo que pondría de manifiesto que aún en estos momentos cualquier discurso histórico artístico e institucional contenido en el medio digital se sitúa casi por

⁴ Mundy, Jennifer y Burton, Jane, “Online Exhibitions”, en Nancy Proctor y Rich Cherry (eds.), *Museums and the Web 2013*, <<http://mw2013.museumsandtheweb.com/paper/online-exhibitions/>> [18/06/2013].

definición en la periferia del discurso museológico más tradicionalista, y como se recalca en esta cita, es todavía necesaria una concienciación sobre su potencial y valor entre casi todas las esferas del conocimiento implicadas y relacionadas con la elaboración y fruición de las exposiciones virtuales.

Retomando la cuestión de la disolución autorial, la exposición no destaca especialmente por incorporar novedades en este sentido, aunque no es ajena a las tendencias de corte colaboracionista y democratizador que los museos están adoptando en los últimos tiempos. En un principio se pretendía complementar el discurso sobre la pérdida con testimonios de visitantes que describieran sus vivencias sobre el argumento, aunque se optó finalmente por no incluirlos. Y la mayor licencia que se ha permitido, para favorecer la pluralidad de voces en la construcción de la narración histórico-artística, ha sido la de elaborar en paralelo un *blog* con entradas escritas por la propia comisaria, o supervisados por ella, con el objeto de garantizar su relevancia académica y curatorial sobre las diferentes obras y artistas que componen la totalidad de la exposición.

El detalle más destacable del *blog* residía en ofrecer un amplio abanico de enlaces de artículos de periódicos, revistas, museos o páginas web correlacionados con el contenido de la exposición (fuentes a las que se confiere de una autoridad similar a la de la Tate); pero especialmente, por facilitar los enlaces a videos de *YouTube* o galerías de *Flickr* cuyos contenidos han sido generados en su mayoría por *amateurs*. Este último detalle posiciona claramente a la institución hacia la apertura, pero al mismo tiempo la sigue situando como un ente revestido de autoridad que controla los contenidos existentes en la red, garantizando con sus elecciones lo que tiene valor o no en Internet.

Aunque el papel del usuario cada vez se está mimetizando más con el del autor, es particularmente interesante abordar e investigar en los procesos y comportamientos del usuario en la red. Precisamente en este sentido ha sido la teoría que estudia la literatura digital la que ha profundizado en cuestiones que nos son de gran utilidad en su aplicación al campo de la museología digital, al extrapolar funciones y conceptos meramente textuales a la combinación de imágenes, textos, videos o sonidos que componen las narrativas hipermediales.

De entre los conceptos que establece Goicoechea (2010), y posteriormente, Sanz y Goicoechea (2012), para describir los procesos y estrategias del usuario para elaborar y construir su propia narrativa se pueden aplicar a este caso de estudio la entropía, la écfrasis y la serendipia.

La entropía entendida como la cantidad de desorden que contiene o libera un sistema, es uno de los conceptos clave para entender la propia idiosincrasia de Internet. Esta provocaría una pérdida de la dimensión temporal y debilitaría la construcción de la narración. En este sentido, interpretando los datos de la media de tiempo por visitante de 6,18 minutos (Mundy y Burton, 2013), esta sería mayor que la de muchas páginas web, pero aún así no es comparable con el tiempo que se requiere para visitar una exposición convencional, lo que nos da una idea de la dificultad para mantener la atención e interés del visitante, no sólo en esta exposición sino en cualquier otra.

La écfrasis entendida según la tradición clásica como el texto que describe una obra visual (Romero, 2012: 313), implicaría en esta situación diferentes medios, y no únicamente la imagen, a través de los cuales el usuario elabora la narración. En el caso de *The Gallery of Lost Art*, de hecho, se ha aportado

un tipo de documentación gráfica, videográfica, sonora o escrita diferente de cada una de las obras. De la capacidad del usuario para contrarrestar tanto los efectos de la entropía además de la variedad de medios que se encuentran en la exposición, depende en gran medida la correcta reelaboración de la narración histórico artística, que en el mejor de los casos, llevaría al descubrimiento serendípico, casual, visto como un elemento sorpresivo y de recompensa al satisfacer la atención prestada por el usuario.

Esta correcta y eficaz adquisición y elaboración de los discursos y narrativas expositivos, no sólo depende de la autoridad museística y curatorial, sino que requiere la determinación y predisposición por parte del usuario para interpretar el contenido en profundidad. El usuario ideal sería aquel denominado por Manovich como *explorador* (2005), y que se contrapone al indolente y superficial *flâneur*; y siguiendo esta misma línea, el usuario ideal para Sanz y Goicoechea "will be one who has acquired the skills to deal successfully with information overload, one who has been able to construct a manageable symbolic object that has found a balance between chaos and order, redundant and relevant information, ambiguity and determined meaning"⁵.

Precisamente en este punto es necesario retomar el papel del usuario como constructor de narrativas que puede llegar a equipararse a la institución museística, como se ha podido ver en la selección de contenidos enlazados en el blog de la exposición. Esta elaboración de contenidos de calidad depende a su vez de una materialización y reelaboración eficaz de las narrativas histórico artísticas. Curiosamente gracias al perfil de *Storify*⁶ se puede valorar la respuesta de los usuarios a diferentes niveles en twitter, y aunque la respuesta de los mismos en las redes sociales no ha sido particularmente participativa (cuestión que daría pie a posteriores estudios acerca del calado de iniciativas digitales abiertas hacia el público), muchas de ellas se podrían interpretar claramente como eventuales micro-narrativas resultantes de la interacción ideal y constructiva que se espera del usuario modelo que sería capaz de aprovechar al máximo cada detalle de cualquier exposición.

BIBLIOGRAFÍA

- Bal, Mieke, "Telling, Showing, Showing off", en *Critical Inquiry*, 18 (1992), Chicago University Press, pp. 556-594.
- Bal, Mieke, *Teoría de la Narrativa. Una introducción a la narratología*, Madrid, Cátedra, D.L. 1995.
- Barthes, Roland, "La muerte de un autor", en *El susurro del lenguaje. Más allá de la palabra y de la escritura*, Barcelona, Paidós, 1994.
- Foucault, Michel, *La arqueología del saber*, México, Siglo XXI, 1970.
- Foucault, Michel, *El orden del discurso*, Barcelona, Tusquets, 2002.
- Foucault, Michel, *Las palabras y las cosas: una arqueología de las ciencias humanas*, México, Siglo XXI, 2006.
- Landow, George P., *Hipertexto: la convergencia de la teoría crítica contemporánea y la tecnología*, Barcelona, Paidós, 1995.

⁵ Sanz, Amelia y Goicoechea, María, "Literary reading rituals and practices on new interfaces", *Literacy and linguistic computing*, vol. 27, Issue 3, September 2012, Oxford University Press, p. 342.

⁶ <<http://storify.com/GalleryLostArt>> [07/07/2013].

- Lyotard, Jean-Francois *La Condición Postmoderna. Informe sobre el saber*, Madrid, Cátedra, 1991.
- Manovich, Lev, *El lenguaje de los nuevos medios: la imagen en la era digital*, Barcelona, Paidós, 2005.
- Mason, Rhiannon, "Cultural theory and Museum Studies", en Sharon Macdonald (ed.), *A companion to museum studies*, Oxford, Blackwell Pub., 2006, pp. 17-32.
- Mundy, Jennifer y Burton, Jane, "Online Exhibitions", en Nancy Proctor y Rich Cherry (eds.), *Museums and the Web 2013*, <<http://mw2013.museumsandtheweb.com/paper/online-exhibitions/>> [18/06/2013].
- Museums in a digital age*, Ross Parry (ed.), London, Routledge, 2010.
- Parry, Ross, *Recoding the museum: digital heritage and the technologies of change*, London, Routledge, 2007.
- Reading moving letters. digital literature in research and teaching: a handbook*, Roberto Simanowski, Jürgen Schäfer y Peter Gendolla (eds.), Bielefeld, Transcript, 2010.
- Romero López, Dolores, "Poetics of crisis or crisis of poetics in digital reading/writing?: The case of Spanish digital literature", *Literacy and linguistic computing*, 27 (2012), Oxford University Press, pp. 305-320.
- Sanz, Amelia y Goicoechea, María, "Literary reading rituals and practices on new interfaces", *Literacy and linguistic computing*, 27 (2012), Oxford University Press, pp. 331-345.
- Teoría del hipertexto*, George P. Landow (ed.), Barcelona, Paidós, 1997.



RESUMEN

El medio digital está permitiendo y favoreciendo un proceso de cambio en los discursos ligados a la institución museística. En paralelo a estas nuevas modalidades discursivas, surgen también tipologías narrativas que reflejan nuevas formas de transmitir y concebir la historia del arte y las propias obras.

La elección de este caso en concreto, la exposición virtual *The Gallery of Lost Art* de la Tate, responde a su relevancia en el marco de las prácticas que se están sucediendo en la actualidad en el ámbito internacional. Si este ejemplo ha supuesto un paso adelante en el diseño de discursos museísticos y exposiciones virtuales, su temática es igualmente novedosa, al realizar un recorrido por obras del arte contemporáneo perdidas bajo diversas circunstancias, que han sido recuperadas y mostradas al público nuevamente gracias a las posibilidades de Internet.

Tomando como punto de partida la noción foucaultiano de discurso, y prestando especial atención a la cuestión de la legitimación institucional de los conceptos que lo estructuran, se pretenden analizar los diferentes elementos que componen el discurso museológico. Así mismo, en este análisis crítico también se hará presente la concepción textológica bajo la que se entiende al museo y que nos llevará a comprender a óptica desde la que se generan las narrativas hipermedia características del medio digital.

Palabras clave: Museología digital, narrativas digitales, nuevos medios.

ABSTRACT

Digital media is encouraging and fostering a changing process of museistic institution discourses. At the same time, these new methods of discourse create new narrative typologies that reflect of new forms of transmitting and conceiving art history and works of art.

I have chosen this particular case, the virtual exhibition Tate's Gallery of Lost Art, because its relevance to current practices that are taking place worldwide. Whether this example has been a step forward in the design of museums' discourses and virtual exhibi-

tions, it theme is just as new. This is due to it being a review of contemporary artworks that were lost under different circumstances but they were been recovered and on display to a large portion of the public thanks to the Internet.

Starting with Foucault's notion of discourse and paying attention specially to matter of the institutional legitimacy of its structural concepts, the aim is to analyze the different elements that make up museum's discourse. Furthermore within this critical analysis I am also considering the textological understanding of a museum, which is leading to better understanding of the point of view being generated from hypermedia narratives.

Keywords: Digital museology, digital narratives, new media.

Primera aproximación para la extracción automática de *Entidades Nombradas* en corpus de documentos medievales castellanos

M^a EUGENIA IGLESIAS MORENO

Depart. Biblioteconomía y Documentación, Depart. Informática, Universidad Carlos III de Madrid
meugenia.iglesias@uc3m.es

PILAR AZCÁRATE AGUILAR-AMAT

Depart. Biblioteconomía y Documentación, Depart. Informática, Universidad Carlos III de Madrid
pilar.azcarate@uc3m.es

SONIA SÁNCHEZ CUADRADO

Depart. Biblioteconomía y Documentación, Depart. Informática, Universidad Carlos III de Madrid
sonia.sanchez.cuadrado@uc3m.es

1. INTRODUCCIÓN

El reconocimiento y clasificación de entidades nombradas (NE – Named Entities)¹ supone una tarea importante para la identificación de nombres propios de personas, lugares y organizaciones. Constituye una subtarea para la extracción y recuperación de información. Los sistemas para el reconocimiento de NE utilizan habitualmente técnicas basadas en métodos estadísticos o en gramáticas lingüísticas. Se han desarrollado diferentes investigaciones para el reconocimiento de NE, sobre todo en inglés aunque en los últimos años se han desarrollado diversos trabajos para otros idiomas como el árabe (Shaalán y Raza, 2007), turco (Metin, Kışla y Karaoğlan, 2012) o español (Galicia-Haro, Gelbukh y Bolshakov, 2004). Para los corpus de documentación medieval los estudios se han centrado, principalmente, en la composición y etiquetado de corpus informatizados con el propósito de elaborar diccionarios y diferentes instrumentos de representación del conocimiento como tesauros u ontologías. Es el caso, a nivel nacional, del *Tesouro Medieval Informatizado da Lingua Galega (TILG)*², el *Corpus Diacrónico del Español (CORDE)*³, *Corpus de Documentos Españoles anteriores a 1700 (CODEA)*⁴, *CHARTA*⁵ y el *Proyecto Biblia medieval*⁶. A nivel internacional mencio-

¹ Chinchor N.: MUC-7 Named Entity Task Definition (1997). <http://www-nlpir.nist.gov/related_projects/muc/proceedings/muc_7_proceedings/overview.html>.

² Proyecto de investigación realizado en el Instituto da Lingua Galega (ILG) de la Universidad de Santiago de Compostela (USC). <<http://ilg.usc.es/tmilg/>>.

³ RAE. <<http://corpus.rae.es/cordenet.html>>.

⁴ Grupo de Investigación de Textos para la Historia del Español (GITHE). Universidad de Alcalá. <<http://demos.bitext.com/codea/>>.

⁵ Corpus Hispánico y Americano en la Red: Textos Antiguos. <<http://www.charta.es/>>.

⁶ <<http://www.bibliamedieval.es/>>.

naremos el *Corpus Informatizado do Português Medieval (CIPM)*⁷, el *Tesoro DEM Informatizado (DEMi)*⁸ y el *corpus Hispanic Seminary of Medieval Studies (HSMS)*⁹.

El objetivo de este trabajo es analizar los problemas surgidos en la anotación automática de los nombres propios contenidos en documentos medievales en castellano medieval con herramientas de procesamiento del lenguaje natural (PLN). Para ello, utilizaremos la aplicación Freeling para español estándar y la variante diacrónica del español de los siglos XII al XVI (Sánchez-Marco, Boleda y Padró, 2011) en un corpus formado por los documentos que componen el *Libro Becerro de las Behetrías de Castilla* (en adelante LBB), del siglo XIV. En esta primera aproximación, nos hemos centrado en la evaluación manual de los resultados obtenidos en el etiquetado de los antropónimos y topónimos. Proponiendo finalmente una adaptación de los módulos de la aplicación con el propósito de aumentar el nivel de acierto del etiquetado y facilitar el posterior reconocimiento y clasificación de las Entidades Nombradas en un trabajo futuro.

Este artículo se organiza en los siguientes epígrafes: en la Sección 2 revisamos las características del corpus LBB. En la Sección 3 describimos el método seguido para la selección del corpus de anotación, evaluación manual de errores y propuesta de adaptación de Freeling. En la Sección 4 se mostrarán los resultados de la evaluación manual. Finalmente, presentaremos algunas conclusiones y trabajo futuro (Sección 5).

2. DESCRIPCIÓN CORPUS LBB

Para esta investigación se ha seleccionado el corpus LBB, compuesto por la transcripción de Gonzalo Martínez Díez (1981) de los 2.109 documentos que integran el *Libro Becerro de las Behetrías de Castilla* de mediados del siglo XIV, en formato electrónico. Es un censo escrito en romance dónde se registran más de dos mil poblaciones castellanas, agrupadas en quince merindades y localizadas en el territorio de Castilla situado al norte del río Duero (Estepa Díez, 2003). El corpus LBB contiene un total de 210.609 tokens, de los que 4.460 son diferentes (Tabla 1).

La elección de este corpus documental se justifica, principalmente, porque posee una alta representación de nombres propios de persona y lugar, objetivo principal de estudio en nuestra investigación.

Tabla 1. Descripción del Corpus LBB

Corpus LBB	
Nº Documentos	2.109
Nº Caracteres	1.085.080
Nº Palabras	208.278
Nº Palabras diferentes	4.268
Nº Tokens	210.609
Nº Tokens diferentes	4.460

⁷ Centro de Linguística de la Universidade Nova de Lisboa (CLUNL). <<http://cipm.fcsh.unl.pt/>>.

⁸ Anterior *Diccionario del Español Medieval (DEM)*. Centro de Investigación de la Academia de Ciencias y Letras de Heidelberg. <<http://www.adw.uni-heidelberg.de/dem/fichero/ficherolista.html>>.

⁹ Proyecto del Dictionary of the Old Spanish Language (DOSL). University of Wisconsin-Madison. <<http://www.hispanicseminary.org/index-es.htm>>.

Los documentos se distribuyen por merindades, división administrativa del reino de Castilla en el siglo XIV, a la que pertenecía cada una de las poblaciones registradas (Tabla 2).

Tabla 2. Distribución de documentos del Corpus LBB

<i>Nº documentos por merindad</i>			
<i>Nº merindad</i>	<i>Nombre merindad</i>	<i>Nº documentos</i>	<i>Nº palabras</i>
I	Cerrato	93	10.854
II	Infantazgo de Valladolid	64	6.848
III	Monzón	93	10.340
IV	Campos	68	5.967
V	Carrión	119	9.579
VI	Villadiego	104	10.007
VII	Aguilar de Campóo	262	23.933
VIII	Liébana-Pernía	129	10.641
IX	Saldaña	194	15.468
X	Asturias de Santillana	179	28.585
XI	Castrojeriz	114	15.436
XII	Candemuñó	73	6.393
XIII	Burgos-Ubierna	117	8.359
XIV	Castilla Vieja	371	34.311
XV	Santo Domingo de Silos	129	11.557
TOTAL:		2.109	208.278

Un análisis manual previo del corpus LBB nos ha permitido determinar su estructura y las principales características respecto a los nombres propios. Teniendo en cuenta que se trata de un censo, el corpus presenta homogeneidad en la estructura de los textos. Los registros de cada población, en la mayoría de los casos, incluyen los siguientes elementos: i) nombre del lugar, ii) obispado al que pertenece (opcional), iii) relación socio-jurídica, acompañada seguidamente del señor (señores) o institución al que pertenece, iv) derechos del rey y v) derechos del señor o institución.

El análisis de las características se ha centrado en la revisión de signos de puntuación, de transcripción, detección de variaciones gráficas y diferentes estructuras de los antropónimos y topónimos propias del castellano medieval. Para evitar problemas en el posterior procesamiento automático de segmentación y etiquetado, se ha procedido a la adaptación manual correspondiente del texto.

Los signos de transcripción incluidos en el texto aparecen entre corchetes o paréntesis y las palabras agregadas en letra cursiva. Se tomaron las siguientes consideraciones:

- Eliminación de los signos que contenían información aclaratoria del transcriptor ([sic], [en blanco], [?]) que no formaban parte del texto original.
- Eliminación de paréntesis y corchetes incorporando los caracteres al texto para un primer análisis.

Los nombres propios dentro del texto se destacan gráficamente con mayúsculas, según indican las normas de transcripción de Gonzalo Martínez Díez (1981: 105), "(...), en la utilización de mayúsculas y minúsculas y en el sistema de puntuación seguimos el criterio actual.". Las variaciones gráficas más relevantes que presentan los nombres propios de persona y de lugar son:

- Uso de caracteres que no existen en español moderno "Ç" y "ç".
- Variaciones del uso de u y v. Por ejemplo, Val Buena de Duero se escribe como "Ual Buena" y "Val Buena".
- Variaciones con omisión o cambio de un carácter para un mismo nombre (Pedro y Pero, Royz y Ruyz, etcétera).

Los antropónimos presentan diferentes estructuras formadas por un elemento (nombre), dos elementos (nombre + apellido patronímico/toponímico) o tres elementos (nombre + apellido patronímico + topónimo). En algunos casos preceden al nombre otros elementos como la fórmula de tratamiento don o donna (don Pedro, don Fernando Sanchez de Ualladolid, donna Margarita, etcétera).

La elisión de una vocal se indica mediante un apóstrofo, *d'* o *D'*. En los nombres propios de persona compuestos se localiza al comienzo de los apellidos para indicar la pertenencia de individuos a familias o poblaciones, siguiendo la fórmula *artículo + d' + familia/lugar* (los d'Aça, los d'Escouadas de Yuso, etc.). En los nombres de lugares es utilizado para diferenciar o situar la población a la que hace referencia (Castriel d'Oniello, Castro d'Ordiales). En el corpus LBB se han localizado un total de 185 nombres propios que contienen apóstrofo.

Los errores tipográficos que han dado lugar a la adaptación correspondiente del texto en formato electrónico se pueden agrupar en:

- Omisión de puntos a final de frase.
- Minúsculas después de punto final que corresponden a punto y coma.
- Apóstrofo en *d'* y *D'*, que indica la elisión de una vocal en los nombres propios, incluido como acento.

3. MÉTODO

El método para la evaluación del etiquetado automático de los nombres propios que aparecen en los documentos del Corpus LBB, se ha diseñado considerando que partimos de un corpus no anotado con las características descritas en el epígrafe anterior. Dicho método es el resultado de la combinación del adoptado en proyectos para enriquecer automáticamente corpus históricos con información lingüística, basado en el etiquetado automático con herramientas existentes seguido de la corrección humana, y el planteado por Sánchez-Marco, Boleda y Padró (2011) para la adaptación de la aplicación Freeling para la variante diacrónica del español de los siglos XII al XVI. Se ha estructurado en las siguientes fases:

- Selección de un corpus de anotación.
- Anotación automática del corpus y evaluación manual de los resultados de etiquetado.
- Propuesta de adaptación de la herramienta de PLN, Freeling.

3.1. Selección de un corpus de anotación

Para la evaluación del etiquetado automático de nombres propios con una herramienta de PLN, en esta primera aproximación, se ha seleccionado una muestra del corpus LBB que hemos denominado corpus de anotación. Este corpus de anotación está formado por 803 documentos, correspondientes a las 7 primeras merindades de las 15 que componen el corpus completo (Cerrato, Infantazgo de Valladolid, Monzón, Campos, Carrión, Villadiego y Aguilar de Compóo) y contienen un total de 78.201 tokens, de los que 2.141 son diferentes. Se localizaron todas las palabras que comenzaban por mayúscula candidatas a ser nombres propios, obteniendo un total de 15.338, siendo 1.250 diferentes (Tabla 3).

Tabla 3. Distribución de tokens en Corpus LBB y corpus anotación

	<i>Corpus LBB</i>	<i>Corpus anotación</i>
Nº tokens	210.609	78.201
Nº tokens diferentes	4.460	2.141
Tokens comienzan mayúsculas	38.369	15.338
Tokens diferentes comienzan mayúsculas	2.739	1.250

3.2. Etiquetado automático del corpus de anotación de LBB

Para el procesamiento automático de los documentos del corpus de anotación, hemos tomado como punto de partida la versión 3.0 de Freeling¹⁰. La elección de FreeLing 3.0 como herramienta de etiquetado se ha basado, principalmente, en que como define Lluís Padró (2011), Freeling es una librería de código abierto para el procesamiento multilingüe automático, que proporciona una amplia gama de servicios de análisis lingüístico para diversos idiomas y, que además, incluye una ampliación para el español antiguo (Sánchez-Marco, Boleda, y Padró, 2011).

La arquitectura modular de FreeLing permite el acceso y modificación de los diferentes recursos lingüísticos que incluye (lexicones, gramáticas, diccionarios, etc.), lo que nos ha permitido extender los datos contenidos en los archivos con vocabulario y formas lingüísticas propias del castellano medieval (siglo XIV), contenidas en el corpus LBB.

El proceso de etiquetado morfológico se ha llevado a cabo en dos iteraciones con la aplicación analyzer incluida por defecto en Freeling 3.0, la primera con los módulos para el español estándar y antiguo facilitados y la segunda con nuestra propuesta de adaptación.

3.2.1. Evaluación manual de errores de etiquetado automático de Nombres Propios

En la primera iteración se realizó un análisis morfológico del corpus de anotación aplicando las herramientas para ambas variantes del español, actual y antiguo, de Freeling. Este analizador está basado en las etiquetas EAGLES (v.2.0)¹¹

¹⁰ Freeling 3.0. <<http://nlp.lsi.upc.edu/freeling/>>.

¹¹ Etiquetas EAGLES. <<http://www.lsi.upc.edu/~nlp/tools/parole-sp.html>>.

que asigna NP00000 a los nombres propios. De los archivos resultado de éste primer análisis se han extraído, mediante scripts en Python, los tokens con la etiqueta NP00000 obteniendo un total de 8.939 para el español estándar y 2.254 en el caso del español antiguo. Tras una primera revisión manual del etiquetado se determinó modificar la configuración de análisis para el español antiguo que contemplara las opción de reconocimiento de entidades nombradas (NE) y de palabras compuestas o *multiword* imprescindible para los nombres propios compuestos. Se obtuvo un resultado de 5.731 tokens etiquetados con NP00000.

Los principales errores detectados han sido los siguientes:

- **Nombres de persona con la estructura *donna* + *Nombre Propio*(NP).** En el texto se incluyen nombres de persona que sólo aparecen bajo la forma *donna* + *NP*, determinante para reconocer las variantes de nombres de persona a lo largo del corpus. El analizador reconoce la palabra *donna* como nombre común femenino singular (NCFS000) y como adjetivo calificativo femenino singular (AQ0FS0). Confirmamos que este problema no surgía con las expresiones de la forma *don* + *Nombre Propio* (NP).
- **Números romanos en mayúsculas.** La finalidad principal de los documentos del corpus LBB es definir a quién correspondía percibir los derechos y rentas expresadas con números romanos en mayúsculas. Esto supone un problema en el análisis con Freeling 3.0 para español estándar que asigna una etiqueta de Nombre Propio en unos casos y de Nombre Común en otro.
- **Nombres propios seguidos de número romanos.** Por ejemplo, *don Nunno VI* o *Diego Garçia los XIII*, en este caso el analizador etiqueta como nombre propio la expresión entera al considerarla una palabra múltiple o *multiword*.
- **Formas verbales propias del castellano medieval.** Por ejemplo: *Pagauan*, *Dixieron*.
- **Artículos, pronombres, conjunciones, contracciones y adverbios propios del español antiguo.** Reconoce como NP determinados artículos, pronombres, adverbios y conjunciones escritos en mayúsculas al comienzo de frase. Por ejemplo: *E*, *Lon*, *Otrosi*, *Deste* (*de* + *este*), etcétera.

3.2.2. Propuesta de adaptación de la herramienta de PLN, Freeling

Nuestra propuesta ha incluido la adaptación de Freeling 3.0 mediante la ampliación de los siguientes módulos:

- Inclusión de *donna* en la lista de palabras que deben formar parte como afijo de las Named Entities en el módulo *Basic NER*, sección *<Affixes>* como "donna PRE". El resultado ha sido el etiquetado correcto de todas las expresiones del corpus de anotación.
- Ampliación del diccionario de español estándar con la lista de números romanos en minúscula, formas verbales, artículos, pronombres, conjunciones y adverbios propios del español antiguo.
- Modificación del submódulo del diccionario dónde se establecen las reglas para sufijos (*<Suffixes>*). Se han sustituido los patrones para que reconozca pronombres enclíticos propios del español antiguo.

Para la segunda iteración, se ha seguido el mismo método que en la anterior, pero utilizando la herramienta con las modificaciones propuestas. Después de la

extracción de los tokens etiquetados con NP00000, se han obteniendo un total de 4.570 tokens para español estándar y 4.485 para el español antiguo.

4. RESULTADOS

La evaluación manual de los resultados de ambas iteraciones nos ha permitido contrastar el nivel de acierto de la anotación automática de nombres propios de personas y lugares con nuestra adaptación de la herramienta Freeling (Tabla 4). En este primer análisis hemos considerado como acierto el reconocimiento y etiquetado completo de todos los elementos que integran la estructura de nombres propios.

En la primera iteración con Freeling 3.0, se obtuvo un nivel de acierto de un 49,21% para español estándar siendo el principal error de etiquetado los números romanos que se escriben con mayúsculas en los documentos del corpus. En este caso suponen un 46,04% de los errores de etiquetado. Para español antiguo (old-es), el nivel de acierto ha sido de 75,41%.

En la segunda iteración con la adaptación propuesta de la herramienta Freeling los niveles de acierto han sido de 97,81% para español estándar, que supone un 48,6% de mejora frente al obtenido en la iteración anterior (Fig. 1), y 98,23% en el caso del español antiguo (Fig. 2).

Tabla 4. Evaluación de resultados de la anotación automática de nombres propios

	1ª Iteración		2ª Iteración	
	FreeLing_Orig_ES	FreeLing_Orig_OLD	FreeLing_Modif_ES	FreeLing_Modif_OLD
NP Total	8939	5731	4570	4485
NP Diferentes	1.769	1.599	1.460	1.403
Aciertos	4.398	4.322	4.470	4.406
Errores	4.541	1.409	100	79

Una vez analizados los resultados de la segunda iteración, podemos concluir que se han solucionado los siguientes errores de etiquetado como nombre propio: i) todos los números romanos del corpus de anotación se le ha asignado una etiqueta de número, ii) se han reconocido las expresiones *donna* + *NP*, iii) formas verbales, artículos, pronombres, conjunciones, contracciones y adverbios propios del español antiguo en el corpus de anotación, se han etiquetado correctamente.

5. CONCLUSIONES Y TRABAJO FUTURO

En este artículo hemos presentado una primera aproximación a la anotación automática de nombres propios en corpus de documentación en castellano medieval. Como primer paso, se ha evaluado el etiquetado de la herramienta de procesamiento del lenguaje natural, Freeling, en dos iteraciones con un corpus seleccionado del corpus LBB: i) Freeling para el español estándar y antiguo por defecto y, ii) Freeling adaptado con nuestra propuesta.

Etiquetado NP Freeling Español

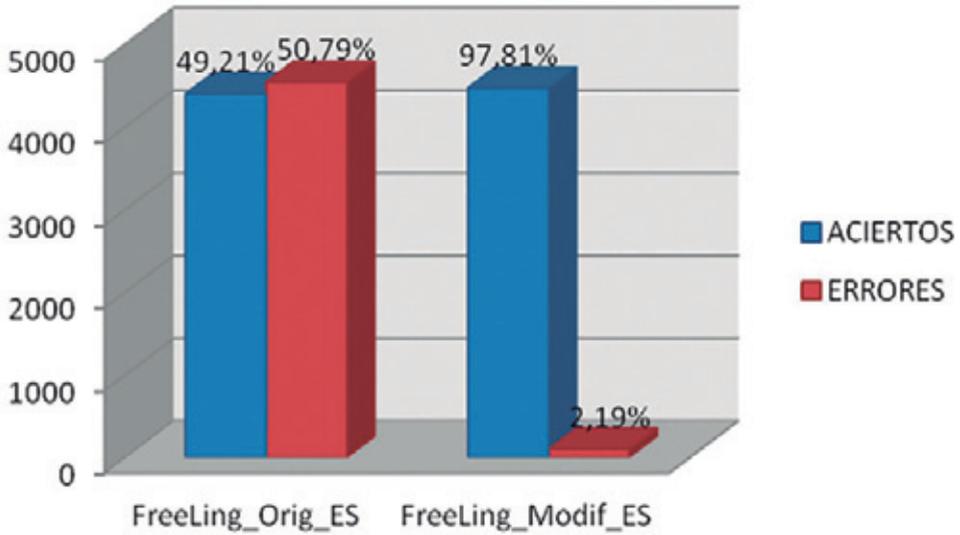


Fig. 1. Nivel de acierto de etiquetado nombres propios con Freeling para español estándar.

Etiquetado NP Freeling Old-Es

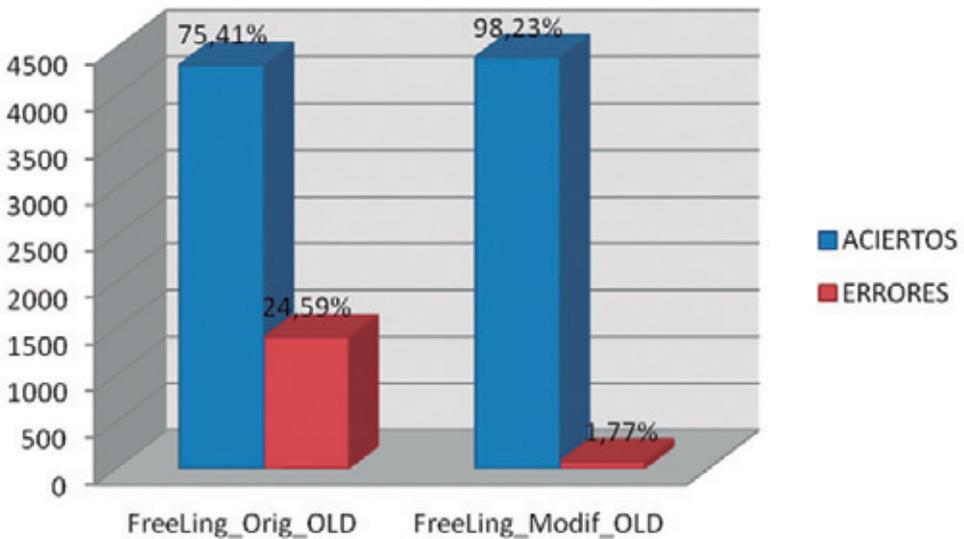


Fig. 2. Nivel de acierto de etiquetado nombres propios con Freeling para español antiguo (old-es).

La revisión manual de los resultados nos ha permitido detectar y clasificar los errores, diseñando nuestra aproximación con el objetivo principal del etiquetado correcto de las estructuras que forman los nombres propios, los números romanos, artículos, pronombres, conjunciones y adverbios, utilizados en el español antiguo basándonos en los que aparecen en el corpus.

Este trabajo, comparando el nivel de acierto de la anotación automática, nos indica que existe un margen para la mejora en los resultados de etiquetado de nombres propios de personas y lugares con una herramienta de procesamiento del lenguaje natural. Así, el nivel de acierto alcanzado en la segunda iteración, 97,81% para el español estándar y 98,23% para español antiguo, nos hace pensar que supone un resultado aceptable para el etiquetado de este tipo de documentos, teniendo como referencia el nivel de precisión de 94,5% obtenido por Sánchez-Marco *et al.* (2011).

Como trabajo futuro nos planteamos repetir el método presentado con el corpus LBB completo con la posterior evaluación para confirmar estos primeros resultados obtenidos. Así como, la extracción y clasificación de las Named Entities de personas, lugares e instituciones del corpus.

AGRADECIMIENTOS

Para este trabajo se ha utilizado la transcripción del *Libro Becerro de las Behetrías de Castilla* de Gonzalo Martínez (1981), contenida en la base de datos cedida por el Grupo de Investigación "QUAESTIO. Sociedades medievales: marcos, redes y procesos", al que pertenecen los doctores Cristina Jular y Julio Escalona, ambos investigadores titulares del Instituto de Historia del CSIC.

BIBLIOGRAFÍA

- Estepa Díez, Carlos, *Las behetrías castellanas*, Valladolid, Junta de Castilla y León, Consejería de Cultura y Turismo, 2003.
- Galicia-Haro, Sofía N.; Gelbukh, Alexander y Bolshakov, Igor A., "Recognition of named entities in spanish texts", en Raúl Monroy *et al.* (eds.), *MICAI 2004: Advances in artificial intelligence, Third Mexican International Conference on Artificial Intelligence*, Mexico City, Mexico, April 26-30, Springer, 2004, pp. 420-429.
- Marrero, Mónica *et al.*, "Named Entity Recognition: Fallacies, Challenges and Opportunities", *Computer Standards & Interfaces*, 35 (2013), pp. 82-89.
- Martínez Díez, Gonzalo, *Libro Becerro de las Behetrías*, León, Centro de Estudios e Investigación San Isidoro, 1981.
- Metin, Senem Kunoba; Kışla, Tarik y Karaoğlan, Bahar, "Named Entity Recognition in Turkish Using Association Measures", *Advanced Computing: an International Journal*, 3 (2012), pp. 43-49.
- Padró, Lluís, "Analizadores multilingües en freeling", *Linguamatica*, 3 (2011), pp. 13-20.
- Padró, Lluís y Stanilovsky, Evgeny, "FreeLing 3.0: Towards Wider Multilinguality". *Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA*, Estambul, European Language Resources Association (ELRA), 2012, pp. 2473-2479.
- Shaanan, Khaled y Raza, Hafsa, "Person name entity recognition for Arabic", *Proceedings of the 2007 Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources*, Praga, Association for Computational Linguistics, 2007, pp. 17-24.
- Sánchez-Marco, Cristina; Boleda, Gema y Padró, Lluís, "Extending the tool, or how to annotate historical language varieties", *Proceedings of the 5th ACL-HLT Workshop on*

Language Technology for Cultural Heritage, Social Sciences, and Humanities, USA, Association for Computational Linguistics, 2011, pp. 1-9.

Xavier, María Francisca, "Variação e mudança lexical no Português Medieval o caso dos verbos", *DOMÍNIOS DE LINGU@GEM – Revista Eletrônica de Lingüística*, 3-2 (2009), pp. 224-242.



RESUMEN

En este artículo presentamos los resultados de una evaluación de la anotación de nombres propios de forma automática en un corpus de documentación medieval castellana. Dicha evaluación se ha realizado sobre el etiquetado obtenido con la herramienta de procesamiento de lenguaje natural, Freeling, en dos iteraciones. La primera, con la versión para español estándar y antiguo facilitadas y la segunda con una adaptación propuesta, basada en la solución de los problemas de anotación debidos a las características y variantes que presentan los nombres propios de personas y lugares en español antiguo. Para ambas iteraciones, se ha seleccionado un corpus de anotación de los documentos que componen el *Libro Becerro de las Behetrías de Castilla* (LBB), del siglo XIV. El nivel de acierto obtenido en la anotación automática de nombres propios con la adaptación propuesta ha sido de 98,23% para el español antiguo, que puede considerarse aceptable para repetir, en un trabajo futuro, el método en el corpus completo.

Palabras clave: Lingüística de Corpus, Anotación de corpus, Documentación medieval, Reconocimiento y Clasificación de Entidades Nombradas.

ABSTRACT

This paper presents the results of evaluating the automatic recognition and annotation of proper names in a corpus of Castilian medieval documents. The evaluation has been done by adapting Feeling, an existing tool for natural language processing. This paper describes the two iterations of this evaluation: the first iteration, using the version for standard and old Spanish, and the second iteration, using an adaptation that has been created based on the problems found in the first iteration. Such problems were mainly caused by the inherent characteristics and variants of proper names and names of places in old Spanish. For that purpose, a corpus of 14th century documents of the *Libro Becerro de las Behetrías de Castilla* (LBB) was used. The proposed adaptation for old Spanish leads to a 98.23% level of success, which indicates that it can be used in the future evaluation of the entire corpus.

Keywords: Natural Language Processing, Medieval Documents, Corpus Annotation, Named Entity Recognition and Classification.

Experiencias de e-research y divulgación digital en historia del arte: GracmonDocs y Viquimodernisme

XURXO INSUA
MIREIA FREIXA
TERESA-M. SALA

Gracmon, Universitat de Barcelona
gracmon@ub.edu

Gracmon es un grupo de investigación consolidado de referencia en el estudio del modernismo catalán, el *noucentisme* y la historia del diseño en Catalunya. Con más de 20 años de actividad, el grupo integra investigadores de los departamentos de Historia del Arte y Diseño e Imagen de la Universitat de Barcelona (www.ub.edu/gracmon).

Desde 2007, con la incorporación de un técnico de apoyo a la investigación especializado, Gracmon adopta metodologías propias del *e-research*, que implican el uso de entornos de trabajo colaborativo y el desarrollo de estrategias de publicación y divulgación *online*.

Esta comunicación pretende dar cuenta, en forma resumida, de dos experiencias en este sentido: *GracmonDocs* –una serie de microsites temáticos que muestran aspectos de la investigación del grupo mediante visualizaciones de datos interactivas– y *Viquimodernisme* –un proyecto Universidad-Wikipedia para la mejora de contenidos en la enciclopedia libre y la redefinición de su uso en contextos académicos.

GRACMONDOCS¹

El primer microsite de la serie es CronoFAD (www.ub.edu/gracmon/docs/cronofad), una cronología *online* que muestra un siglo de historia del Foment de les Arts i del Disseny, institución fundada en 1903 en Barcelona para la promoción de las artes decorativas y que a lo largo de su trayectoria ha sido un agente central en el apoyo a la cultura del diseño.

Esta institución fue el objeto de estudio del proyecto de investigación financiado dentro del Plan Nacional de I+D+i *El Fomento de las Artes Decorativas de*

¹ www.ub.edu/gracmon/docs [fecha de consulta para todas las url citadas: 01/10/2013].

Barcelona (1903-1992) (BHA2003-03215), dirigido por Mireia Freixa. A partir de sus resultados, y dentro del proyecto *El Sistema Diseño Barcelona: visualización y genealogía histórica* (HUM2006-052252/ARTE), a cargo de Anna Calvera, se elaboró este microsite.

CronoFAD está técnicamente basada en *Timeline*², un widget AJAX de código abierto para la visualización de datos temporales desarrollado inicialmente por David François Huynh y el grupo SIMILE del MIT (Massachusetts Institute of Technology). Es relativamente sencillo de incrustar y personalizar, y se puede enlazar a bases de datos muy fáciles de crear.

En el caso de CronoFAD, las modificaciones en la implementación se centraron en dos aspectos: el diseño de una interfaz mejorada, y la construcción de un back-end para la introducción dinámica y colaborativa de los datos. Este segundo aspecto se solucionó ágilmente con Django, gracias a la participación de un programador externo, que dispuso un panel de administración gráfico en el que los eventos a mostrar en la cronología pueden ser añadidos por diversos usuarios autorizados de forma simple. En cuanto a la interfaz, y basándose en criterios de usabilidad, coherencia y claridad, se trató de rediseñar la apariencia por defecto del widget para acercarla a las necesidades específicas del proyecto y hacerla, en general, más atrayente y sencilla.

Así, organizados de forma sintética, CronoFAD muestra más de un millar de eventos pertenecientes a todas las áreas de actividad de la institución, así como sus órganos de dirección, los locales donde se alojó con su ubicación, los diversos premios convocados y un largo etcétera. Todos estos acontecimientos aparecen dispuestos sobre el fondo de sucesivas etapas históricas generales identificadas como fundamentales e incluidas a modo de contexto.

El usuario puede moverse adelante y atrás en el tiempo arrastrando la línea temporal o saltando directamente a la etapa de su elección, acceder a cada evento obteniendo información complementaria y/o imágenes y realizar búsquedas mediante la introducción de texto. Esta búsqueda puede hacerse *filtrando* los resultados, de forma que únicamente se muestren en la cronología los eventos que contengan el texto introducido (p. ej. "exposición") o bien *resaltándolos* con distintos colores, lo que permite realizar búsquedas combinadas relativamente más complejas (obteniendo como resultado, por ejemplo, las exposiciones realizadas durante el mandato de un determinado presidente).

La siguiente entrega de GracmonDocs se desarrolló al hilo del proyecto financiado *Los espacios del ocio en la Barcelona de 1900. Un proyecto de e-research en torno a los orígenes de la industria cultural* (HAR2008-4327/ARTE), dirigido por Teresa-M. Sala. El proyecto, que fue el primero en articularse desde el inicio en torno a metodologías *e-research*, se propuso cartografiar los lugares de divertimento y distracción en el cambio de siglo modernista. Se identificaron centenares (casi ochocientos) cinematógrafos, teatros, salas de baile y otros espacios para la música, galerías de arte y estudios de fotógrafos, con sus respectivas ubicaciones, textos descriptivos e imágenes de época.

La visualización *Mapa cronológico de los espacios de ocio en la Barcelona del 1900* (<www.ub.edu/gracmon/docs/mapaeob>) muestra estos espacios sobre la trama urbana de la ciudad. Un manejador nos permite movernos en el tiempo,

² <www.simile-widgets.org/timeline>.

entre 1870 y 1910, de manera que podemos seguir su desarrollo diacrónico, observando cómo se van clausurando o desplazando locales, y abriendo otros nuevos, y con ello no obtener sólo una vista de detalle, sino visualizar panorámicamente los grandes cambios en los ejes del ocio barcelonés que se producen en la época. Paralelamente, puede observarse la evolución urbanística de la propia ciudad: los espacios de ocio están ubicados sobre sucesivos mapas de época, proporcionados por el Institut Cartogràfic de Catalunya, que van cambiando también a medida que nos desplazamos en el tiempo.

El *Mapa cronológico...* está construido usando únicamente herramientas gratuitas, como la aplicación Google Earth, su plug-in para navegadores y las hojas de cálculo en la nube de Google Spreadsheets. Concretamente, se trata de una implementación de la iniciativa Google Outreach, que ofrece tutoriales y plantillas al alcance de un usuario básico.

Mientras que el plug-in de Google Earth nos permite ver el *Mapa* desde el propio navegador, la aplicación en sí nos da la posibilidad de, una vez descargado el archivo .kmz correspondiente que se ofrece en la web, disponer al lado del mapa de un índice completo de los locales ordenados por tipo y realizar búsquedas de texto.

La base de datos de la que se nutre la visualización es una simple hoja de Google Spreadsheets, lo que permite una edición colaborativa sumamente ágil y simplificada (los datos fueron introducidos por más de una decena de investigadores).

Dentro del mismo proyecto financiado sobre *Los espacios de ocio...* se desarrolló una línea de trabajo sobre los antecedentes de las actividades de recreo en el período estudiado. Así, se trabajó sobre el siglo XVIII a través de un estudio de *Calaix de Sastre*, el dietario que el Barón de Maldá escribió entre finales del XVIII e inicios del XIX. Esta obra es uno de los textos más relevantes de la narrativa en catalán previa a la Renaixença, y tiene una importancia capital como fuente de conocimiento sobre el ocio de la época. Se extrajo de él una selección de pasatiempos y diversiones considerados significativos y la visualización *Calaix de Sastre, Eventos de ocio en la Barcelona del XVIII* (<www.ub.edu/gracmon/docs/calaix>) los muestra ordenados en una línea de tiempo.

Entre las opciones que ofrece esta visualización está la posibilidad de acceder a los comentarios del Barón acerca de los eventos señalados, extraídos en forma literal del texto original. Se incluye también una función de búsqueda por texto, en la que introduciendo cualquier cadena de texto se devuelven los eventos que lo contengan. Los eventos, además están categorizados por su carácter público o particular, para la clase social de sus participantes y por el tipo de distracción, de manera que podemos hacer un filtrado facetado de los que más nos interesen. Por ejemplo, si queremos saber cómo eran las diversiones públicas de la menestralía que consigna el Barón en su diario, sólo tenemos que hacer la selección oportuna y los eventos mostrados se reducirán a sólo aquellos que cumplan ambos requisitos.

En el momento de presentar esta comunicación, diversos *GracmonDocs* se hallan en proceso. Mencionar además la línea de trabajo recién abierta, en colaboración con el Departamento de Matemática Aplicada de la Universitat de Barcelona, para su rediseño en forma de app para dispositivos móviles o la generación de otros nuevos y específicos para estas plataformas.

VIQUIMODERNISME³

A modo de introducción mínima: Wikipedia es una enciclopedia *online* libre, gratuita y editada colaborativamente por miles de usuarios voluntarios, disponible en más de 280 idiomas, cuyos contenidos se ofrecen bajo licencia Creative Commons BY SA, y entre cuyas normas de funcionamiento está la necesidad de citar fuentes (Wikipedia no es una fuente primaria).

Wikipedia recibe unos 20.000 millones de visitas al mes, siendo el sexto site más visitado en internet. Su relevancia en la práctica como puerta de acceso cotidiana al conocimiento básico es difícil de soslayar, sobre todo teniendo en cuenta que una gran mayoría de búsquedas en Google devuelven como primer resultado el artículo correspondiente en Wikipedia.

Desde Wikipedia y la Wikimedia Foundation (la organización *non-profit* que le da soporte) se puso en marcha en años recientes la GLAM-Wiki Initiative, GLAM siendo el acrónimo inglés de "galerías, bibliotecas, archivos y museos", para colaborar con estas instituciones en su labor de difusión cultural mediante proyectos colaborativos con la comunidad wikipedista. Entre las entidades participantes en diversos proyectos están el British Museum o la Bibliothèque nationale de France, y en nuestro entorno más próximo, el Museu Picasso, MNAC...

En este contexto, Gracmon se asocia con Amical Wikimedia, asociación sin ánimo de lucro para la promoción de la Wikipedia en catalán (Viquipèdia), para llevar a cabo un proyecto innovador a nivel global, involucrando a los docentes-investigadores del grupo, alumnos de grado y postgrado de Historia del Arte y la comunidad de editores de la Wikipedia.

Como grupo de investigación especializado en *modernisme*, el objetivo planteado fue el de mejorar la calidad de los contenidos relativos al modernismo catalán presentes en Viquipèdia, teniendo en cuenta su importancia como vehículo para la divulgación y la relevancia patrimonial del *modernisme* en Catalunya.

En primer lugar, durante el verano de 2012 se llevó a cabo una auditoría previa para determinar el estado de dichos contenidos (que se compararía *a posteriori* con los resultados finales del proyecto, permitiendo un seguimiento riguroso de su impacto). Los indicadores a partir de los cuales se vertebraría la auditoría fueron el número de artículos en las categorías relacionadas con el modernismo catalán, número de bytes del artículo, número de imágenes cargadas en Wikimedia Commons sobre un elemento o categoría, número de imágenes utilizadas para ilustrar cada artículo, número de enlaces internos y número de visualizaciones (variables cuantitativas); distinción (o no) como artículo de calidad, disponibilidad y número de traducciones, uso y cantidad de referencias y utilización de categorías para agrupar y ordenar los artículos (variables cualitativas).

La auditoría está disponible al completo en Viquipèdia⁴. En breve, señalar la existencia de 372 artículos sobre modernismo catalán (relativamente escasos en relación a la cobertura de temas similares), con una extensión muy desigual, y

³ <<http://ca.wikipedia.org/wiki/Viquiprojecte:Viquimodernisme>>. Esta parte de la comunicación se basa en la presentada por Àlex Hinojo y Esther Solé (impulsores del proyecto) en el congreso internacional *coupDefouet* celebrado en Barcelona en junio de 2013 (actas en prensa).

⁴ <https://ca.wikipedia.org/wiki/Viquiprojecte:Gracmon/Auditoria_prèvia>.

de los que únicamente 26 tenían la distinción de artículos de calidad. Un artículo típico tendría 45 kb de datos, tres traducciones y un número de referencias (una clave fundamental de calidad) extremadamente variable (entre 0 y 122). Son cifras relativamente pobres, pero aún así el conjunto de artículos recibía del orden de 50.000 visitas al mes (cifra que ascendía a 600.000 si se contaban los artículos correspondientes en otros idiomas). La auditoría, en conclusión, mostró en datos en qué medida los contenidos sobre *modernisme* precisaban una mejora sustancial tanto en términos cualitativos como cuantitativos, y apuntó posibles líneas de acción, que fueron luego concretadas y articuladas en torno al wikiproyecto *Viquimodernisme*.

Para ponerlo en marcha, antes del inicio del curso se celebró una reunión entre voluntarios de Amical Wikimedia y Gracmon. Se elaboró una lista de artículos a mejorar, que se distribuyeron entre las asignaturas (de grado y posgrado) elegidas para participar en el proyecto en el primer trimestre, con docencia a cargo de profesores integrantes de Gracmon: *Barcelona 1900*, *Fuentes de la Historia del Arte* y *De las Artes decorativas en la cultura del diseño*.

La propuesta para los estudiantes se hizo en forma de redefinición del trabajo académico típico en relación al conocimiento libre *online*. En lugar de plantear un trabajo genérico básico, sin otro recorrido que la evaluación por parte del docente, en el que es lógicamente inaceptable "copiar de Wikipedia", se invitó a los alumnos a escoger uno de los artículos propuestos para ampliarlo o crearlo en Wikipedia.

Dependiendo de cada asignatura y su estructura de evaluación preexistente, el trabajo se realizó individual o colectivamente y tuvo carácter obligatorio o voluntario, formando parte necesariamente de la nota final o como complemento para la mejora de la calificación. El tiempo disponible para su realización era el cuatrimestre al completo e incluía una presentación oral una vez terminado. Para que los alumnos obtuvieran el conocimiento necesario para empezar a editar en Wikipedia, se organizaron talleres presenciales, a cargo de miembros de Amical Wikimedia.

Así, las tareas de búsqueda bibliográfica, síntesis y redacción se adaptaron al formato wiki, añadiéndose además imágenes tomadas de Wikimedia Commons o realizadas por los propios autores y subidas a dicho repositorio. Se recomendó ir actuando directamente sobre la página de la entrada elegida en Viquipèdia, para trabajar colaborativamente con el resto de la comunidad desde el primer momento, disponer de una plataforma común y transparente para la comunicación entre alumnos, profesores y wikipedistas, e ir solucionando los posibles problemas. Estos, si bien menores, de hecho se dieron, fundamentalmente en dos sentidos.

En primer lugar, cuando los alumnos veían cómo, por ejemplo, se eliminaban referencias erróneas que ellos habían incorporado (otros wikipedistas las habían comprobado) o se corregían directamente cuestiones de formato. El hecho de que el control sobre los contenidos era necesariamente compartido con otros, con una comunidad más amplia, generó tensiones en un primer momento, que se resolvieron una vez aclarado el funcionamiento de Wikipedia y observando cómo en la práctica la dinámica conducía a artículos de mayor calidad. Señalar, por otro lado, que a la hora de la evaluación por parte de los docentes, fue útil la página de historial con la que cuentan todos los artículos de Wikipedia, que permiten seguir la evolución en las ediciones y determinar con precisión quién

las realiza (y así, por ejemplo, comprobar el nivel de implicación de los distintos miembros en los trabajos en grupo).

En segundo lugar, la falta de conocimiento detectada en torno a temas relacionados con los derechos de autor propició una cierta incompreensión inicial cuando se comprobó que no cualquier imagen podía ser utilizada para ilustrar una entrada. Wikimedia Commons (el repositorio de contenido audiovisual de Wikipedia) sólo admite contenidos bajo licencia CC BY SA o menos restrictiva (como las de dominio público). Además, el propio sistema de subida de imágenes a Commons, que podría mejorarse, ofreció numerosas dudas técnicas, solucionadas con otro taller presencial *ad hoc*.

En el segundo cuatrimestre se contó con la asignatura *Corrientes estéticas del teatro del siglo XX*, donde los estudiantes realizaron biografías de dramaturgos, siguiendo los parámetros del cuatrimestre anterior. El trabajo fue evaluable y voluntario, y también se realizó una clase presencial para formar a los alumnos en edición wiki.

En total, en ambas fases participaron 111 estudiantes, mejorando 57 artículos y creando 38 nuevos, llevando al artículo típico de la auditoría previa a una media de 7 traducciones y 25 referencias. 9 artículos más están en proceso de cumplir los restrictivos criterios que distinguen a un artículo de calidad⁵. El incremento en contenido, respecto al estadio previo, puede cifrarse en torno al 500%.

En general, se han cubierto deficiencias notorias y mejorado la calidad en numerosos conceptos, destacando por ejemplo los artículos relacionados con publicaciones periódicas aparecidas en época modernista.

Respecto a la satisfacción del alumnado con la iniciativa, los resultados de la encuesta realizada muestran, por ejemplo, que un 16% de los participantes editó más artículos de los que les correspondía y un 73% afirma que volvería a repetir la experiencia, demostrando un potencial motivador importante en propuestas de este tipo. Esta motivación se deriva también de la percepción de utilidad del trabajo hecho, que pasa a convertirse en un contenido abierto y disponible, en proceso continuo de ampliación y traducción, muy fácilmente accesible, y en definitiva, socialmente útil.

CONCLUSIÓN

Estas dos experiencias son una muestra proactiva de las implicaciones de las tecnologías de la información y la comunicación en el entorno de la Historia del Arte, y exponen en la práctica el hecho de que la innovación en el ámbito académico, en este sentido, es más fácil cuando se colabora con ámbitos externos punteros en conocimiento libre (el caso de Wikipedia).

Por otra parte, señalar a propósito de GracmonDocs cómo la adopción desde la investigación académica de herramientas de visualización accesibles de forma abierta y gratuita, al alcance de cualquier usuario y sencillas de implementar y usar, suponen un avance sustancial frente a proyectos costosos y frecuentemente menos interoperables. La inversión en diseño de interfaz, a menudo pasada por alto, sí puede ser, sin embargo, clave para la obtención de mejores resultados

⁵ <http://ca.wikipedia.org/wiki/Viquipèdia:Què_és_un_article_de_qualitat>.

en cuanto a usabilidad, y apunta a la importancia de incluir diseñadores en los equipos multidisciplinares característicos de las humanidades digitales.



RESUMEN

Esta comunicación tiene por objeto la presentación de experiencias y proyectos de Gracmon (Grupo de Investigación en Historia del Arte y del Diseño Contemporáneos, Universitat de Barcelona) en el entorno de las Humanidades Digitales y la transferencia del conocimiento en internet. Se discutirán cuestiones metodológicas y técnicas, con un enfoque fundamentalmente descriptivo y práctico.

Desde hace 5 años, Gracmon (cuyos principales objetos de estudio son el modernismo, el *noucentisme* y la historia del diseño en Cataluña) viene desarrollando su compromiso con el *e-research* a través de iniciativas de trabajo colaborativo, visualización de datos y divulgación *online*. Presentaremos dos propuestas: *GracmonDocs* y *Viquimodernisme*.

GracmonDocs son una serie de micrositos monográficos, cada uno de los cuales presenta un tema concreto de la investigación del grupo mediante una visualización interactiva. Las visualizaciones aportan una visión sintética y unas posibilidades de búsqueda y filtrado que las convierten en herramientas útiles en sí mismas para el investigador. Asimismo, por su accesibilidad y sencillez de uso, adquieren un carácter divulgativo.

En cuanto a *Viquimodernisme* (Wikimodernismo), se trata de un proyecto innovador de colaboración de una institución académica con Wikipedia. La iniciativa se propone la mejora y ampliación de los contenidos sobre modernismo en la Wikipedia en catalán. Para ello, se involucra a investigadores, docentes y estudiantes redefiniendo la relación del trabajo académico con Wikipedia.

Palabras clave: e-research, visualización de datos, Wikipedia, GLAM Wiki, modernisme.

ABSTRACT

This paper presents some experiences and projects by Gracmon (Research Unit on the History of Contemporary Art and Design, University of Barcelona) in the field of Digital Humanities and knowledge transfer on the internet. Methodological and technical issues will be discussed in a fundamentally descriptive and practical way.

For 5 years, Gracmon (whose main objects of study are *modernisme*, *noucentisme* and the history of design in Catalonia) has been developing a commitment to e-research through collaborative online work, data visualization and online outreach initiatives. Two proposals are presented: *GracmonDocs* and *Viquimodernisme*.

GracmonDocs are a series of monographic microsites, each of which showcases a specific topic of the group's research through an interactive data visualization. They provide a panoramic view of the subject for the general public, while at the same time, via search and filtering options, they can be useful to the researcher.

Viquimodernisme constitutes an unprecedented collaboration between an academic institution and Wikipedia. This GLAM wikiproject seeks to improve and expand the contents of *modernisme* in the Catalan Wikipedia (*Viquipèdia*). Researchers, professors and graduate and postgraduate students are involved, redefining the relationship of the academic assignment with Wikipedia.

Keywords: e-research, data visualization, Wikipedia, GLAM Wiki, Catalan Art Nouveau.

Los clubes del libro en la era digital: nuevas estrategias y aspectos sociales. El caso de Círculo de Lectores (1998-2012)

RAQUEL JIMENO REVILLA
Centro de Ciencias Humanas y Sociales (CSIC)
raquel.jimeno@cchs.csic.es

INTRODUCCIÓN

Desde su creación a comienzos del siglo XX, los clubes del libro con fines comerciales han supuesto una fuente de innovación en estrategias empresariales y de difusión de la lectura dentro del mundo editorial. La búsqueda de un nuevo mercado y de la oferta de un producto diferente a los que podían encontrarse a través de medios tradicionales les ha conducido tanto a las innovaciones gráficas y artísticas dentro del libro en su concepción de objeto artístico, como al descubrimiento de infraestructuras empresariales que les permiten llegar a un público no habitual, obteniendo como resultado la consiguiente creación de nuevos lectores.

Hay que tener en cuenta a la hora de valorar estas aportaciones, en primer lugar, la innovación que supone dentro del mundo de las técnicas editoriales y comerciales la creación y difusión de los clubes del libro ya que, como indica Pascal Fouché (1998), se pasa del acto consciente y voluntario de un cliente de comprar un libro previamente elegido, a otro que implica una pasividad mucho mayor de recibirlo como consecuencia de un acuerdo previo (y de una oferta previamente seleccionada) o, incluso, de un olvido, en el caso de los clubes que contemplan la opción negativa (envío automático de una obra en caso de que no se rechace de antemano).

De esta manera, los clubes del libro rompen el esquema tradicional de la mediación editorial, al tratar de lograr el vínculo directo con el socio suprimiendo la cadena de intermediarios comerciales, con el fin de conseguir una relación directa entre el nivel de la selección/edición de textos y el de la recepción de los mismos. Con ello se pretende lograr beneficios a partir de ofrecer una mayor comodidad al cliente, tanto en la selección como en el acceso al texto. Esta ruptura conlleva un cambio en las estrategias comerciales, promocionales y logísticas, de cuyas innovaciones se beneficiará también la totalidad de la industria editorial.

El sector de lectores al que se dirigen específicamente estos clubes se encuentra, bien por motivos geográficos, bien por su nivel de alfabetización y cul-

RAQUEL JIMENO REVILLA

tura, en una situación de difícil acceso a las librerías, por lo que los clubes del libro cumplirían una función de creación de público lector. Por ello comenzaron a establecerse precisamente en aquellos lugares y épocas en los que se daba un importante déficit de librerías (Estados Unidos a comienzos del siglo XX, Alemania tras la Primera Guerra Mundial o España bajo el franquismo son algunos ejemplos). Paulatinamente, el aumento de lectores logrado por los clubes del libro, que también repercute en las librerías, así como el consecuente beneficio en todos los eslabones de la industria editorial (autores, editores, distribuidores, libreros) y la comprobación de que la existencia simultánea de ambas formas de venta es posible (llegando a tener varios clubes –como France Loisirs o el propio Círculo de Lectores– sucursales o puntos de venta físicos propios), ha logrado la aceptación de los clubes que, en palabras de Fuinca (1993: 136):

Los clubs del libro se configuran "incubadoras" de lectores, y que después de una permanencia media de 2 a 3 años, el socio tiende a abandonar el club y realizar sus compras por otros cauces: librerías, distribución a gran escala, etcétera. Además, los clubs estimulan las compras de libros exteriores al club; por ejemplo, en Francia se adquieren por cada 6 libros dentro del club, 5 fuera; en Italia, compraban libros exclusivamente en el club el 72% de los socios en 1971 y sólo el 55% de los socios en 1991.

Asimismo, las propias librerías se benefician de las innovaciones estratégicas de los clubes, adaptándolas a su funcionamiento y logrando con ello un mayor beneficio. Como afirman Huret y Cerisier (2007: 13):

L'édition traditionnelle, même si elle continuera à se vivre essentiellement dans la méditation (la presse, la librairie, la prescription), s'appropriera pourtant une part non négligeable de ces pratiques, notamment pour la publication de séries volumineuses ou pour la promotion de ses collections (ventes à prime périodiques). Les supports de communication directe de l'éditeur seront, les années passant, plus nombreux. L'apport des clubs est également perceptible en librairie; face à cette nouvelle concurrence, qui redéfinissait la pratique du conseil au lecteur, les points de vente traditionnels n'eurent d'autre choix que de repenser leurs propres méthodes de fidélisation de la clientèle.

De esta forma, su labor no es excluyente con las librerías, sino complementaria, puesto que éstas también se benefician del aumento del público lector (al no permanecer muchos de los nuevos lectores únicamente en el ámbito del club), así como de la adaptación de las estrategias comerciales creadas por los clubes para su propio beneficio.

En resumen, el objetivo fundamental de los clubes del libro es la captación y la fidelización del cliente, tanto mediante el enriquecimiento de la edición original de un texto (que provocará el importante aporte de la innovación creativa al grafismo dentro del mundo de la edición) como con el desarrollo de una serie de estrategias orientadas a la consecución de descuentos respecto a la edición en librerías y la obtención de gratificaciones por permanecer en el club. El desarrollo de dichas estrategias, así como la búsqueda y creación de público lector será diferente en el caso concreto de cada club. Estudiaremos el caso concreto de Círculo de Lectores, con el fin de ver en qué medida dicho club ha empleado los nuevos recursos digitales para lograr sus fines, adaptándose a la evolución de un público lector que, sobre todo en las últimas décadas, ha visto influenciados por los avances tecnológicos tanto sus formas de ocio como las preferencias a la hora de elegir sus lecturas.

EL CASO DE CÍRCULO DE LECTORES

La creación de Círculo de Lectores en 1962 no es resultado de la casualidad. España vive en ese momento una serie de cambios fundamentales que van a modificar profundamente las bases políticas, económicas y sociales del país que había sido construido después del triunfo del régimen franquista. Este clima de mayor apertura y desarrollo del turismo fomenta la llegada de empresas extranjeras (es este el caso del grupo alemán Bertelsmann). La lenta aparición en España de una clase media que aspiraba a mejorar su nivel de vida, también en el plano cultural, pero que no estaba habituada a frecuentar librerías ni bibliotecas, es aprovechada por Círculo para desarrollarse a partir de la oferta de una selección de textos en ediciones de calidad, que eran llevadas directamente al domicilio del socio.

En lo que concierne a las características concretas, es necesario comenzar por destacar el funcionamiento de Círculo puerta a puerta, característica que lo diferencia de la práctica totalidad de clubes del libro a nivel mundial, y que es consecuencia del deficitario sistema de correos que había en España en el momento de la fundación del club. Este factor es el punto de partida de una infraestructura que permite el contacto directo con el cliente, ya que el agente está encargado tanto de repartir la revista con el catálogo ofertado cada bimestre, como de recoger el pedido y entregárselo al socio, una vez que es recibido. De esta manera, Círculo obtiene, mediante la relación y el consejo directo del agente al socio, la creación de un vínculo emocional y una garantía mayor de la fidelización de este último, que puede confiar en un agente que conoce sus gustos y preferencias personales. Partiendo de este punto, Círculo desarrollará estrategias de adaptación para obtener una fidelización lo más larga posible.

Los primeros años de existencia de Círculo no fueron fáciles. El verdadero despegue sucedió con la llegada del que será su directivo más importante, Hans Meinke que, en su primera etapa en el club como asistente del director Arnold Schmitt entre 1965 y 1968, obtiene un resultado de medio millón de socios para el club. Los años setenta se caracterizaron por la diversificación de la oferta de productos, comenzando a incluir en el catálogo toda clase de objetos que nada tienen que ver con los libros. Además, en la oferta literaria, Círculo acentuó la tendencia al *best-seller*, en lugar del equilibrio anterior entre el *best-seller* y la literatura de calidad. Este error estratégico estuvo a punto de acabar con la empresa. Es por ello que un nuevo cambio de dirección culminó con el retorno de Hans Meinke a comienzos de los años ochenta, quien lograría a lo largo de diecisiete años el aumento del número de socios hasta un millón y medio.

Meinke fundamentó el giro cultural del club en tres líneas: el retorno al equilibrio entre la literatura de calidad y el *best-seller*, la atención al cuidado en las ediciones y, por último, el comienzo de una actividad cultural paralela con la creación de la Fundación Círculo de Lectores, encargada del fomento de actividades como ciclos de conferencias, encuentros con los escritores o exposiciones de artistas destacados. En este punto es necesario remarcar asimismo la labor de Círculo como editor, que culmina con la creación de la editorial Galaxia-Gutenberg, presente en las librerías. Asimismo, Círculo crea su filial catalana Cercle de Lectors y Círculo del Arte, con el mismo sistema de funcionamiento que el club original pero dedicado a la obra gráfica de artistas plásticos, que en su mayor parte ya habían ilustrado libros para el club.

Con la jubilación de Hans Meinke en 1997, su sucesor Albert Pélach continuó la misma línea de trabajo, incorporando también los avances de las nuevas tecnologías, como la creación de la página web, que sirve también para realizar el pedido de productos. Los cambios en Bertelsmann a nivel europeo culminaron en 2010 con la venta del 50% de las acciones de Círculo al grupo editorial Planeta, uno de los grupos de mayor proyección en España e Hispanoamérica. La unión de estos dos grupos ha posibilitado iniciativas como la plataforma digital Booquo, presentada en febrero de 2012, que pretende convertirse en un sistema pionero en la distribución de libro electrónico y cine en España.

LAS PÁGINAS WEB DE CÍRCULO DE LECTORES

En su labor de adaptación a las nuevas tecnologías, Círculo de Lectores inauguró su página web en 1997, como afirma Alfred Comín (2001: 17):

Círculo comenzó su andadura en Internet en 1997 con una página institucional. De forma progresiva, el departamento informático de la compañía ha ido elevando el listón. Se publica el catálogo de productos, después asoman con timidez algunos contenidos y con el tiempo llega la posibilidad de encargar pedidos. La necesidad de anticiparse a los cambios del mercado es lo que ha movido ahora a los responsables de Círculo a profundizar su puesta por Internet.

Lo que comenzó siendo una manera de proyección y difusión en su forma originaria de <www.circulodelectores.com>, fue incorporando paulatinamente más funciones, bajo la nueva dirección de <www.circulo.es> (manteniendo <www.circulodelectores.es> y <www.circulodelectores.com> para las ofertas por hacerse socio), de manera que se acaba transformando no sólo en una posibilidad de visualizar el catálogo abierto a todos los internautas, sino también en un método de solicitud de pedidos alternativo a la revista, en la que además pueden consultarse contenidos que, por coste o por formato, no pueden incluirse en el formato papel.

A través de una clasificación de los diversos productos mediante pestañas (libros, música, cine, hogar, club y belleza), los socios registrados pueden valorar y comentar cada uno de los títulos, además de ofrecerse la opción abierta "me gusta" de Facebook o la recomendación a un amigo. Asimismo, se muestra una lista de recomendaciones de títulos afines al seleccionado, de manera similar a otras webs de venta de libros y discos como Amazon, Fnac o la Casa del Libro. En caso de que el artículo seleccionado se desee comprar, existe la opción de incluirlo en el carrito de compra, o bien comprarlo en un clic. Sólo se puede efectuar la compra en caso de ser socio.

Otra de las posibilidades que ofrece la web es un sistema de compra híbrido, mediante la pestaña "Compra rápida" de la parte superior de la página de inicio, que permite introducir todos los artículos que se desean solicitar mediante el código con el que aparecen en la revista, permitiendo de esta manera no sólo una mayor rapidez, sino que aquellos socios que prefieran hojear la revista para elegir su pedido puedan luego contar con la ventaja de la rapidez que supone la solicitud por Internet. Esta diversificación se da también en cuanto a la forma de pago, pudiendo elegir, además de las maneras tradicionales de contra reembolso por correo y a través del agente, la tarjeta de crédito o bien la domiciliación bancaria. De esta forma se facilita también la labor del agente, que no

tiene que estar tan pendiente como antes de localizar al socio en su domicilio para poder recoger la hoja de pedido.

Aparte de las secciones tradicionales dedicadas a libros y música, la web cuenta en su catálogo con una extensión a la oferta de otros productos diversos que se alejan de la oferta inicial de Círculo: cine, hogar y belleza. Todos los productos cuentan con la misma presentación, dividiéndose cada sección en apartados similares para mostrarlos. Para todas las secciones se tiene muy en cuenta la valoración de los socios, ya que aquellos productos más vendidos y valorados ocupan un lugar destacado, junto con las novedades, en el primer puesto de cada sección. Asimismo, se ha creado un apartado titulado "Te damos más" en las secciones de libros, música y cine, dedicado a los contenidos extras para cada tipo de productos, en el que se muestran concursos, noticias y reportajes que pretenden conseguir que los socios se sientan partícipes de los logros y propuestas del club al que pertenecen, así como de enriquecer la oferta de productos con información de la actualidad cultural.

En cuanto a la sección titulada "Club", se alberga una miscelánea de noticias y eventos relacionados con Círculo: desde el Cuestionario Círculo (entrevista formulada a un personaje conocido) y los concursos para los socios, hasta las ofertas especiales, las noticias y los acontecimientos relacionados con el club. También dedica un apartado a esbozar una breve historia del club, que enlaza con la web específica del 50 aniversario; con la recién creada editorial digital Arroba Books, de la que hablaremos más adelante, así como con el Premio de Novela convocado por Círculo anualmente, en el que los socios pueden participar como jurado. Los visitantes de la web también pueden estar al tanto de las ofertas de trabajo relacionadas con Círculo a través del apartado correspondiente. Por último, se detallan las ventajas de ser socio, así como de los privilegios que se consiguen mediante las tarjetas concedidas por años de permanencia en el club.

Como recursos web complementarios, Círculo cuenta con una página propia dedicada al 50 aniversario, conmemorado en 2012 (www.nosquedamuchoporleer.com). Se trata de un espacio virtual dedicado, fundamentalmente, a ofrecer una agenda (sección mostrada en la página de inicio) con los principales actos y eventos en relación con el aniversario (obras de teatro, exposiciones, charlas, entrevistas y apariciones en los medios, etc.). Ofrece, además, en la sección "Nuestra historia", una línea del tiempo con los hitos cronológicos más importantes de cada año, tanto a nivel de Círculo como a nivel español e internacional, complementada con documentos gráficos y visuales, así como un dossier con los anuncios de Círculo ("Círculo en imágenes") que aparecieron en televisión a lo largo de esos años. Dentro de la programación del 50 aniversario se encuentran iniciativas como la realización de ediciones especiales (en torno al número 50), felicitaciones de personalidades conocidas, información sobre el ya mencionado premio de novela y el enlace a un video conmemorativo, en "Mucho por leer", junto con campañas que apelan a la participación del socio, como proyectos sociales con libros ("enLibrate") y un reto solidario; concursos ("quiz literario") y una sección dedicada a las ventajas de la promoción por amistad.

En página web aparte, a la que puede accederse también en la parte inferior de la web principal, "Tu Círculo para papás" (www.tucirculoparapapas.com) presenta, a su vez, concursos y novedades, recomienda lecturas para los más pequeños y oferta planes en familia con descuento. Además, puede consultarse el blog de Tu Círculo (<tucirculo.circulo.es>), otro espacio independiente (con en-

lace a través de la web principal) que, consciente de la atracción que suponen las nuevas tecnologías para el público más joven, adopta el formato de blog para la presentación de sus contenidos, que consisten fundamentalmente en las novedades de literatura, dvd y juegos infantiles y juveniles del catálogo de Círculo, junto con aquellas noticias relacionadas con el club y la posibilidad, en la sección "Diviértete", de descargar juegos, chistes y cómics relacionados con las novedades del catálogo. El alta de socio se realiza a través de los padres, debiendo estos ser socios de Círculo para que pueda efectuarse.

OTROS RECURSOS DIGITALES

La apuesta de Círculo de Lectores por las posibilidades que ofrece la edición digital se refleja en dos vertientes: la creación de la plataforma digital Booquo y la editorial digital Arrobabooks.

Con el lanzamiento de Booquo en febrero de 2012, se presenta un modelo que pretende ser una alternativa a la compra de libro electrónico y películas, basándose no sólo en dicha opción de compra, sino también en la suscripción, a modo de otras páginas de películas o del sistema Spotify en el ámbito musical, siendo accesible además a través de todos los dispositivos. Es necesario un registro gratuito en la página para poder comprar eBooks o alquilar películas, así como para dejar comentarios. También existe la posibilidad de hacerse socio Premium por la que, mediante el pago de una tarifa plana de 9,90 € al mes, se permite el acceso a todos los eBooks (lectura en la nube, préstamo indefinido mientras se mantengan las cuotas) y películas que, en caso de no ser socio Premium, sólo pueden alquilarse por un máximo de 72 horas. Además, hay ediciones de eBooks a las que únicamente se puede acceder con la opción Premium, así como la existencia de canales temáticos con contenidos adicionales, como una revista especializada o un blog de actualidad. Asimismo, se posibilita contactar con el anfitrión del canal, que es un escritor relevante dentro de ese ámbito temático. También se posibilita el acceso a una serie de revistas del grupo editorial, como *Ser Padres*, *Mía* o *Geo*.

Por otro lado, se le concede gran importancia a la creación de una comunidad virtual en la que los usuarios, al registrarse, pueden perfilar sus gustos mediante la realización de un test, editar su perfil, gestionar la cuenta de sus productos adquiridos, acceder a su biblioteca y videoteca, escribir reseñas de un eBook o película, contactar con otros usuarios afines, participar en grupos de discusión y crear los suyos propios o participar en creaciones conjuntas, todo ello con el fin de enriquecer y compartir la experiencia lectora, como se expresa en la web.

Arrobabooks, lanzada a comienzos de este mismo año, es el equivalente en formato eBook a la iniciativa tomada en 1994 con la creación de Galaxia Gutenberg: la distribución de las ediciones exclusivas del club al mercado en general, y no únicamente a los socios. En la sección "Sobre Arrobabooks" de su página web se explican estas características, así como la intención de incorporar 150 nuevos eBooks cada año (comenzando por 50 títulos en honor al aniversario) y la apuesta por formatos singulares como la entrega en serie (heredero de la novela por entregas), el relato corto o los contenidos adicionales. Se declara la intención de mantener los precios entre 0,99 y 9,99 € por cada eBook, y, finalmente, la decisión no exenta de polémica de no incorporar el DRM (*digital Rights Management*) siempre que el contrato editorial lo permita, con el fin tanto

de agilizar la compra como de que los eBooks sean aptos para todos los dispositivos, entendiendo que el DRM no es una medida eficaz contra la piratería. Por el momento, el catálogo disponible se restringe a narrativa, separada por géneros, y a los llamados libros prácticos. Cada título se acompaña de un breve resumen del contenido, con posibilidad también de buscar por autores. La página web está destinada únicamente a catálogo, con la función de enlace a aquellas librerías y plataformas digitales donde está disponible el ejemplar seleccionado (Booquo, Nubico, Fnac, Casa del Libro o Amazon), en formato ePub o Mobipocket.

PÚBLICO LECTOR Y RECEPCIÓN

Como se ha visto en la historia de Círculo resumida en la introducción, España ha atravesado, desde el momento de la fundación del club en 1962, una etapa llena de cambios tan numerosos como trascendentes para su historia, economía, sociedad y cultura. Lógicamente, el público lector al que se dirige Círculo ha evolucionado en la misma medida, influido tanto por los acontecimientos históricos nacionales como por las tendencias sociales y culturales a nivel global, condicionadas fundamentalmente por el ascenso y el predominio de los medios de masas durante esta época. Si bien una de las aportaciones fundamentales de Círculo de Lectores ha sido la creación de un público lector entre un sector de la población sin posibilidad o costumbre de acceso a librerías y bibliotecas, el club ha debido adaptar sus recursos a la sociedad a lo largo de estos años, como hemos podido ver a lo largo de la exposición, con el fin de adaptarse a un público cada vez más familiarizado con las nuevas tecnologías y con las nuevas formas de sociabilidad e intercambio cultural que favorece Internet. Se ha dado, además, una evolución en los intereses de dicho público, que ha pasado de entender la lectura fundamentalmente como un medio de perfeccionamiento personal capaz de otorgar un mayor número de oportunidades para la promoción económica y social, a un concepto de la lectura más en la línea de la cultura de ocio y un enriquecimiento personal a nivel individual, acorde con la sociedad de consumo que comienza a desarrollarse precisamente a partir de los años iniciales del club.

Asimismo, y ante el surgimiento de nuevas formas de ocio como la televisión y los videojuegos, los índices de lectura han descendido en proporción al nivel total de la población, siendo la capacidad y hábitos lectores cada vez menores a pesar del aumento demográfico y la expansión de la educación. A este hecho se le une el surgimiento de la lectura en formato digital, que ha cambiado la relación física con el libro y los condicionantes del mismo por los de los dispositivos empleados, abriendo la posibilidad de una simultaneidad de producción, transmisión y lectura de un mismo texto al poder aunar en un mismo individuo las tareas de escritura, edición y distribución, por lo que se anulan aquellas distinciones que separaban los cometidos intelectuales y las funciones sociales, obligando así a redefinir todas las categorías que hasta ahora formaban parte de lo esperado y percibido por los lectores. Estos dos factores han ocasionado que Círculo de Lectores deba conocer y emplear los nuevos recursos en beneficio propio, no viéndolos como una competencia sino como una nueva posibilidad de promoción y difusión. La creación de su editorial digital y de la plataforma Booquo, junto con las posibilidades de uso y participación que se ofrecen con la página web, así lo demuestran.

CONCLUSIONES

Como se ha podido comprobar a lo largo de la exposición, los recursos que posibilita Internet han tratado de ser empleadas por parte de Círculo de Lectores como un recurso complementario a aquellos que han posibilitado su éxito –relación directa con el socio y creación de un vínculo más emocional que le permite el funcionamiento puerta a puerta–. A un empleo inicial de la página web como un medio de difusión y promoción, se le han ido añadiendo progresivamente funciones de catálogo, medio de interacción entre los socios y el club, así como medio alternativo de solicitud y pago del pedido. Todo ello con el fin de ofrecer al socio tanto una mayor facilidad para realizar la elección del catálogo y la compra de artículos, como una mayor posibilidad de participación a través de los concursos, charlas, noticias, blogs y la posibilidad de realizar sus propios comentarios. De esta manera se persigue el objetivo de hacer sentir al socio partícipe de la comunidad de lectores que pretende crear el club y que, con ello, aumente su fidelización.

Además, Círculo está realizando un esfuerzo por extender su oferta a las nuevas formas de lectura digital que cuentan con una difusión y acogida cada vez mayor, con el fin de seguir manteniendo a aquellos socios que eligen realizar una parte o la totalidad de sus lecturas, películas y música en formato digital, así como de atraer a un público ya familiarizado con estos formatos, ofreciendo una alternativa a los elevados precios del eBook mediante la oferta de una tarifa plana, junto con el empleo de recursos similares a los del club tradicional para difundir un sentido de comunidad de lectores. Asimismo, Círculo se propone continuar con su labor de editor, iniciativa que le ha distinguido del resto de clubes similares de otros países, abriéndose camino también dentro de la edición digital. El cambio de hábitos y preferencias del público, muy acentuado a lo largo de los cincuenta años de existencia de Círculo, ha supuesto la adaptación a una sociedad que ha pasado de ver el libro como una oportunidad de promoción social a través de la formación personal, a una alternativa más de ocio, en competencia con otros medios de más reciente desarrollo (radio, televisión, videojuegos, Internet...). Este hecho implica una continua adaptación a los nuevos recursos digitales con el fin de presentar una alternativa en igualdad de condiciones a las de los dichos medios de ocio, tecnológicamente más avanzados y llamativos.

BIBLIOGRAFÍA

- Barnet, Richard J. y Cavan, John, *Sueños globales. Multinacionales y el nuevo orden mundial*, Barcelona, Flor del Viento, 1995.
- Bloss, Carlos, *Círculo de Lectores. La pasión por el libro. Crónica y balance de un club cultural*, Barcelona, Círculo de Lectores, 1990.
- Burgos, Martine; Evans, Christophe y Buch, Esteban, *Sociabilités du livre et communautés de lecteurs, trois études sur la sociabilité du livre*, París, BPI, 1996.
- Bustamante, Enrique (coord.), *Comunicación y cultura en la era digital: industrias, mercados y diversidad en España*, Barcelona, Gedisa, 2002.
- Cerisier, Alban, "Les clubs de livres dans l'édition française (1946-1970)", *Bibliothèque de l'École des Chartes*, 155 (1997), pp. 691-714.
- Cerisier, Alban, "D'un club à l'autre. Deux générations de clubs de livres en France", *Entreprises et histoire*, 24 (2000), pp. 21-42.

- Comín, Alfred, "Club del libro 2.0", *Windows 2000 magazine*, 53 (2001), pp. 16-17.
- Collins, Jim, *Bring on the books for everybody. How literary culture became popular culture*, Durham and London, Duke University Press, 2010.
- Ferreira, Lola, "Las cuadraturas del Círculo", *Delibros*, 12 (1989), pp. 27-35.
- Fouché, Pascal, *L'édition française depuis 1945*, París, Éditions du Cercle de la Librairie, 1998.
- Frattini, Eric y Colias, Yolanda, *Tiburones de la comunicación. Grandes líderes de los grupos multimedia*, Madrid, Ediciones Pirámide, 1996.
- FUINCA (Fundación para el Fomento de la Información Automatizada), *El sector del libro en España. Situación y líneas de futuro*, Madrid, Fundesco, 1993.
- Hernández, Álvaro, "El club del Dr. Meinke", *Mercado*, 338 (1998), pp. 20-26.
- Huret, Jean y Cerisier, Alban, *Le Club du Meilleur Livre 1952-1963*, París, Librairie Jean Huret, 2007.
- Juristo, Juan, "Hans Meinke: treinta años de Círculo de Lectores", *Delibros*, 47 (1992), pp. 54-55.
- López de Abiada, José Manuel, *Entre el ocio y el negocio. Industria editorial y literatura en la España de los 90*, Madrid, Verbum, 2001.
- Moret, Xavier, *Tiempo de editores. Historia de la edición en España*, Barcelona, Destino, 2002.
- Sánchez, Pedro, *Los clubes del libro en el mundo editorial: el caso de Círculo de Lectores*, <<http://eprints.rclis.org/16330/1/Clubos-Circulo%20Lectores.pdf>> [30/06/2013].
- Vila-Sanjuán, Sergio, *Pasando página: autores y editores en la España democrática*, Barcelona, Destino, 2003.

PÁGINAS WEB

- Los libros más vendidos y las novedades de música y cine. En: <www.circulo.es> [consulta: 05/07/2013].
- Hazte socio de Círculo de Lectores y aprovecha ventajas del club. En: <www.circulodelectores.es> y <www.circulodelectores.com> [consulta: 04/07/2013].
- 50 Aniversario. En: <www.nosquedamuchoporleer.com/Paginas/default.aspx> [consulta: 03/07/2013].
- Tu Círculo para Papás. En: <www.tucirculoparapapas.com> [consulta: 03/07/2013].
- Tu Círculo. En: <tucirculo.circulo.es> [consulta: 04/07/2013].
- Comprar, leer eBooks en tu librería online. En: <www.booqu.com> [consulta: 04/07/2013].
- Arrob@Books. En: <www.arrobabooks.com> [consulta: 03/07/2013].



RESUMEN

Desde su creación a comienzos del siglo XX, los clubes del libro con fines comerciales han supuesto una fuente de innovación en estrategias empresariales y de difusión de la lectura dentro del mundo editorial. La búsqueda de una nueva demanda y de la oferta de un producto diferente a los que podían encontrarse a través de medios tradicionales les ha conducido tanto a las innovaciones gráficas y artísticas dentro del libro en su concepción de objeto artístico, como al descubrimiento de infraestructuras empresariales que les permiten llegar a un público no habitual, obteniendo como resultado la consiguiente creación de nuevos lectores.

En esta ponencia se pretende abordar, centrándose en el caso español de Círculo de Lectores, cómo se han adaptado estas empresas al aprovechamiento de nuevos sistemas de información y comunicación durante las últimas décadas, especialmente en el caso de

Internet. Con este fin, se analizarán las diferencias a nivel de funcionamiento empresarial que trae consigo el hecho de que la relación entre club y socio se gestione digitalmente. Asimismo, se analizarán las nuevas herramientas digitales de las que hacen uso los clubes en la actualidad junto con las finalidades que dichos clubes persiguen con la creación y mantenimiento de las mismas. Por último, se tendrá en cuenta los aspectos sociales de los puntos discutidos anteriormente, centrándonos sobre todo en varios aspectos: el cambio de perfil del público de este tipo de clubes, resultado de la evolución de la mentalidad de dicho público acorde con la época.

Palabras clave: Club del libro, Círculo de Lectores, Internet, edición digital.

ABSTRACT

Since its creation in the early twentieth century, book clubs with commercial purposes have meant a source of innovation in business strategies and diffusion of reading in publishing. The search for a new demand and supply of a product different from those that could be found through traditional means has led them to both graphic and artistic innovations in the conception of the book as an art object, and also the discovery of entrepreneurial infrastructures that allow them to reach an unusual audience, having a subsequent creation of new readers as a result.

This paper seeks to approach, focusing on the Spanish case of Círculo de Lectores, how these companies have been adapted to the use of new information and communication systems in recent decades, especially in the case of the Internet. To this end, we will analyze the differences at the level of business operation that entails the fact that the relationship between club and membership is managed digitally. We will also analyze the new digital tools that book clubs make use of currently, together with those ends which are pursued by clubs with the creation and maintenance of the same. Finally, we will take into account the social aspects of the points discussed above, with special focus on several aspects: changing audience profile of this type of clubs, as a result of the change in attitudes of the public in line with the times.

Keywords: Book club, Círculo de Lectores, Internet, digital edition.

Publicidad digital

SONIA LAVANDEIRA ROJO
Universidade da Coruña
sonialavandeira@gmail.com

El fenómeno publicitario nos rodea en nuestra vida cotidiana. Estamos tan acostumbrados a encender la televisión/la radio, o abrir una revista/periódico y encontrarnos con anuncios de todo tipo, que en la mayoría de los casos ignoramos, o mejor dicho, creemos que ignoramos, pero en realidad sus imágenes impactantes, sus pegadizos eslóganes y sus estrategias lingüísticas son captadas inconscientemente y pueden llegar a formar parte de nuestro lenguaje cotidiano.

Cada año las agencias de publicidad ganan millones por manipular a los posibles consumidores para conseguir su único objetivo: que compren sus productos. Debido a la enorme competencia entre agencias, el anuncio debe ser lo más original posible para que no pase desapercibido. Para ello cuentan con un sinfín de recursos, como por ejemplo, el uso de llamativos eslóganes, uso de imágenes impactantes, la aparición de personas famosas promocionando el producto u ofreciendo muestras del mismo. La publicidad no es un fenómeno moderno. Sus orígenes se remontan a la antigüedad y Russell & Lane (1994: 4) afirman que el primer testimonio de la publicidad son unas tablillas de arcilla encontradas en Babilonia del año 3000 a.C., que incluyen las inscripciones de un comerciante, un copista y un zapatero. También existen papiros que atestiguan que los egipcios usaban ya la publicidad, al igual que los griegos y los romanos, como así lo atestiguan los muros pintados en Pompeya anunciando la venta de carne y leche. Pero el texto oral existió antes que el escrito, por lo que los pregoneros públicos de las civilizaciones hebrea, griega y romana fueron los primeros publicistas (Brewster *et al.*, 1997: 22). Cuando se inventó la imprenta en el siglo XV empezaron a aparecer los primeros anuncios escritos y con la Revolución Industrial en el siglo XVIII se consolidó. Las primeras cuñas publicitarias emitidas por la radio fueron alrededor del año 1922 y los primeros spots televisivos fueron en 1939 (Brewster *et al.*, 1997: 341). Internet se creó en 1960, aunque no fue hasta la década de los noventa cuando se empezó a utilizar como medio público como lo conocemos hoy, y la primera publicidad digital data aproximadamente del año 1992. Actualmente los publicistas deben decidir en qué medio publicitario quieren colocar su anuncio. Siguiendo las indicaciones del lingüista Baños *et al.* (2009) podemos ofrecer la siguiente clasificación de los medios publicitarios. En primer lugar, diferenciaremos medios convencionales y medios no-convencionales. Entre los primeros encontramos

medios como la prensa, la radio, televisión, cine, exterior (como por ejemplo la publicidad en vallas) e internet. Entre los medios no-convencionales podemos nombrar la publicidad directa, como por ejemplo en el reparto de folletos, o la publicidad en el lugar de venta, cuando por ejemplo se nos ofrece una muestra o pequeña degustación de un producto en un mostrador en un centro comercial. Entre los nuevos medios publicitarios, Baños *et al.* (2009) nombra la televisión interactiva y el uso del móvil. Es curioso que un fenómeno tan revolucionario como internet esté mencionado entre los medios convencionales y no en los nuevos medios. Una posible explicación podría estar en que la publicidad llegó con mucho impulso y a estas alturas las nuevas estrategias para sorprender ya estén agotadas.

Los lingüistas Vestergaard y Schroder (1992: 14) definen la publicidad como un proceso comunicativo verbal o no-verbal, público y unidireccional. En efecto, la publicidad es una forma de comunicación donde encontramos todos los elementos que forman el proceso comunicativo. Tenemos el emisor (el anunciante), el receptor (el posible consumidor), el mensaje (el propio anuncio), el lenguaje del discurso publicitario en ocasiones acompañado de imágenes (el código) y el medio publicitario donde aparece el anuncio (el canal). Sin embargo con la llegada de la publicidad digital ya no podemos estar de acuerdo con la definición de Vestergaard y Schroder ya que la nueva herramienta no es unidireccional si no que nos ofrece una interactividad constante que analizaremos a continuación. Veamos antes la clasificación de los diferentes tipos de anuncios que podemos encontrarnos en cualquier medio publicitario. Vestergaard y Schroder (1992: 1-3) distinguen dos tipos de anuncios, los **no-comerciales** (por ejemplo, los panfletos políticos o los de asociaciones benéficas, que no tratan de vender nada) y los **comerciales**. Dentro de los comerciales diferencian tres tipos, los **anuncios de grandes firmas** que anuncian el funcionamiento de la empresa a sus clientes para así recordar el nombre de la firma y dejar una impresión favorable, los **anuncios industriales** donde una firma anuncia sus servicios a otras firmas, estos anuncios suelen encontrarse en periódicos y revistas de economía y los **anuncios de consumo** que pueden ser privados (los típicos anuncios por palabras que no tratan de persuadir al posible comprador) o pueden ser creados por agencias de publicidad. La mayoría de ellos anuncian bienes materiales como comida, bebida y ropa, o bien bienes sociales como por ejemplo perfume o coches.

Como hemos visto los anuncios pueden aparecer en diferentes medios pero aquí nos centraremos en los que aparecen en internet y son estos anuncios a los que nos referiremos como publicidad digital. Podemos encontrarnos portales web dedicados enteramente a anuncios como por ejemplo <adflip.com>, donde encontramos muchos spots anunciando todo tipo de productos como coches, ropa, relojes, comida y cosméticos entre otros. Se puede buscar anuncios por años de publicación desde el año 1940 al 2001, o por el nombre de la revista o periódico donde aparecieron. En este portal también se incluye un *ranking* y dan la oportunidad de votar por tu anuncio favorito. El internauta interactúa y lo más importante para el anunciante, está viendo los anuncios de marcas que en la mayoría de los casos existen a día de hoy y esto puede convertirlos en posibles consumidores del producto por lo que se traduce en más ventas. Hoy en día, la publicidad digital incluye juegos, tests, blogs y sitios interactivos y lúdicos. La publicista Williams (2010) recoge varios ejemplos de diferentes campañas en internet que revolucionaron la publicidad en este medio, como por ejemplo, el

sitio web de *Burger King* o el de *Nike*. En el primero, aparecía un hombre disfrazado de pollo al que los internautas podían dar órdenes como el que bailase o diese una voltereta, aunque todas estas instrucciones estuviesen grabadas el simple hecho de que el "pollo" respondiese lo hacía atractivo. Ahora nos puede resultar una página muy simple, pero tenemos que tener en cuenta que estamos hablando de los primeros anuncios digitales y en su tiempo supuso toda una revolución, y que a partir de ellos se trabajó en métodos más elaborados. En 2007, *Burger King* con motivo del estreno de la película de los *Simpson* lanzó una página web llamada <simpsonizeme.com>, donde los usuarios subían fotos para caracterizarlas con sus personajes favoritos de la serie. Nike también creó con éxito varias campañas publicitarias digitales, entre ellas podemos destacar *The chain* en el que el futbolista Ronaldinho lanzaba un balón y los espectadores tenían que seguir la cadena de lanzamientos, la única condición era que el balón tenía que entrar por la izquierda y salir por la derecha. La participación fue masiva y se creó un video de dos horas de duración con algunos de los videos recibidos. Esta campaña también se reforzó con la aparición del anuncio en la televisión donde se pedía que se viese la página web para poder participar.

Lo novedoso de la publicidad digital es que los consumidores pueden interactuar dando opiniones de los productos en foros, blogs, o en páginas de consumidores como *ciao*, *trnd* o *bopki*. Las redes sociales como *Facebook* contienen anuncios y las diferentes marcas de productos crean grupos de seguidores. Una herramienta como *YouTube* también es usada para alabar, criticar e incluso parodiar alguna campaña publicitaria, o simplemente cuando quieres ver algún contenido tienes que ver antes un anuncio, que a veces puedes omitir tras verlos unos segundos. Un ejemplo del uso de *YouTube* y *Facebook* usados conjuntamente para un mayor impacto sería el anuncio del *Gorilla* para anunciar la marca de chocolate *Cadbury Dairy Milk* recogido en Williams (2010: 73). Este anuncio empezó siendo un anuncio para la televisión, pero alcanzó el éxito total cuando se trasladó como video en *YouTube* recibiendo millones de visitas y a partir de ahí se crearon más de setenta grupos de seguidores en *Facebook*. En el anuncio se veía a un hombre disfrazado de gorila tocando la batería mientras cantaba una canción de Phil Collins. Pronto aparecieron nuevas versiones del anuncio en *YouTube*.

Según el publicista Ocaña (2006) podemos encontrar diferentes formas de anuncio en la red como el *banner*, *skyscraper*, *pop up*, *pop under* y *layer*. Un anuncio en forma horizontal es un *banner*, que si los usuarios de Internet están interesados en ver sólo tienen que cliquear y así obtener más información. Un *banner* también puede resultar ser un *layer*: "Elemento publicitario que aparece sobre la página con animación y movimientos predeterminados" (Ocaña, 2006: 113), porque poco a poco las palabras del anuncio pueden ir cambiando, como así puede suceder en un *skyscraper* (anuncio en forma vertical). A veces, cuando cerramos una página, puede aparecer un cuadro con un anuncio (*pop under*), lo contrario del *pop up*, si el anuncio aparece antes de la página de Internet.

Portales como *Google* ofrecen servicio publicitario llamado *Google AdWords* y *Google AdSense*. Puede resultar cansino navegar por la red y ser bombardeados con *banners*, *pop ups*, *pop unders* y similares. Por lo tanto, se crearon servicios como los que ofrece *Google*. Si estamos, por ejemplo, buscando información sobre "viajes" en el buscador, al margen derecho aparece diferentes links de vuelos baratos o agencias de viajes, cuando pulsamos en cualquiera de ellos, es cuando el anunciante paga por el anuncio. Este método se llama *Google AdWords*.

Sin embargo, cuando el anuncio es incluido en la propia página web, es lo que se denomina *Google AdSense*, y es el dueño de la página quien obtiene beneficio cuando alguien pulsa algún anuncio.

Algunos publicistas como Williams consideran que el salto definitivo a la publicidad *online* aparece con lo que denomina como *publicidad viral*, que define como un anuncio, normalmente en forma de video, que se emite por Internet y va extendiéndose como un virus entre los internautas. El público busca estos videos que pasa del boca oreja, al contrario de las formas de publicidad antes vistas por Ocaña que podrían llegar a irritar y distraer al internauta. La publicidad digital es eficaz en el sentido que capta el interés del público y esto se traduce en más ventas, y lo hace a nivel mundial.

El lenguaje también tiene que adaptarse a las nuevas tecnologías, y el lenguaje publicitario no es ninguna excepción. El lingüista Crystal (2003) llama al lenguaje usado en Internet *Netspeak*, y como tal tiene sus propias características: como el típico léxico del "*Netspeak*", enunciados cortos, uso del presente y párrafos cortos compuestos de oraciones simples. Veamos estas características en el anuncio de Ilvico que apareció en la prensa escrita en el año 2010 donde podemos encontrar el típico léxico como por ejemplo: "alerta", "virus", "detectar", "cancelar" y "aceptar", preguntas como "¿Desea aliviarlos ahora?", enunciados cortos como "ilvico" y debajo "comprimidos recubiertos", el uso de tiempos presentes como en "tómalo" o "no necesita", párrafos cortos compuestos por oraciones simples y a veces en forma de columna, en este caso donde se describe las propiedades de las pastillas como "práctico" y "cómodo". El anuncio en sí también parece una página web donde se incluye la barra de herramientas y el mensaje de alerta que en ocasiones aparece cuando algo va mal. Vemos la imagen de un chico, quizás porque los receptores de este anuncio sea gente joven que está acostumbrada a navegar por la red y está acostumbrada al uso del *Netspeak*, y el anuncio afirma que las pastillas son "compatibles con tu estilo de vida". Internet también está presente en anuncios impresos en prensa escrita, no sólo cuando utilizan el lenguaje propio de la red, sino también cuando incluye las direcciones de Internet del producto o servicio anunciado para que el lector obtenga más información sobre lo anunciado.

Entre las ventajas de la publicidad digital podemos destacar la forma, la sencillez, la rapidez, la interactividad, el bajo coste, el alcance a nivel mundial y los millones de seguidores que tiene la red. Entre las desventajas mencionaríamos que no todas las personas tienen acceso a la red y que los recursos y estrategias publicitarias van quedando obsoletas a gran velocidad, por lo que concluiremos con la reflexión si la publicidad digital debe ya innovarse, porque se puede repetir técnicas pero resultan cansinas y hay que trabajar en nuevos métodos que desbanquen a los anteriores.

BIBLIOGRAFÍA

- Baños González, Miguel; García García, Francisco y Ramírez Perdiguero, Francisco Javier, *Las palabras en la publicidad: el redactor publicitario y su papel en la comunicación publicitaria*, España, Laberinto, 2009.
- Brewster, Arthur Judson; Palmer, H. Hall y Ingraham, R. George, *Introducción a la publicidad*, México, CECSA, 1997.
- Crystal, David, *The Cambridge encyclopedia of the English language*, Cambridge, Cambridge University Press, 2003.



ALERTA VIRUS

SE HAN DETECTADO SÍNTOMAS DE **GRIPE Y RESFRIADO**.
¿DESEA ALIVIARLOS AHORA?

CANCELAR ACEPTAR

ilvico
COMPRIMIDOS RECUBIERTOS

ilvico
comprimidos recubiertos
Activo sintomático de los
síntomas de gripe y resfriado
Via oral
20 comprimidos

Compatible con tu estilo de vida

PRÁCTICO → tómalo donde y cuando lo necesites

CÓMODO → no necesita preparación

ilvico, práctico para aliviar los síntomas de la gripe y el resfriado.

ACTIVALO

MERCK

DE VENTA EN FARMACIAS. Lea las instrucciones de este medicamento y consulte al farmacéutico. No administrar en caso de insuficiencia hepática. No consumir bebidas alcohólicas, no conducir ni realizar actividades peligrosas mientras consume este medicamento. Puede producir somnolencia. CPS M-07514. MERCK FARMA Y QUÍMICA S.L. Polígono Merck - 06100 - Mollet del Vallès - BARCELONA.

Fig. 1. Anuncio Ilvico.

Ocaña, Fernando, *La publicidad contada con sencillez*, Madrid, Maeva, 2006.
Russel, J. Thomas y Lane, W. Ronald, *Kleppner publicidad*, México, Prentice Hall, 1994.
Vestergaard, Torben y Schrode, Kim, *The language of advertising*, Oxford, Blackwell, 1992.
Williams, Eliza, *La nueva publicidad: las mejores campañas*, Barcelona, Gustavo Gili, 2010.



RESUMEN

La publicidad no es un fenómeno actual. Sus orígenes se remontan a la antigüedad cuando empezaron a existir los primeros pregoneros anunciando productos. Sin embargo, con el paso de los años, la publicidad se fue modernizando hasta llegar a la era digital, donde por supuesto encontramos anuncios y publicidad viral. En este trabajo analizamos cómo se introdujeron los primeros anuncios en la red y cómo evolucionaron a día de hoy con la ayuda de *Facebook* y *YouTube*. Ofrecemos también una clasificación de tipos de publicidad y medios dónde pueden aparecer. La publicidad digital es un proceso comunicativo que utiliza estrategias lingüísticas y como tal no podemos olvidar la influencia que ejerce y recibe de la publicidad convencional escrita.

Palabras clave: Publicidad, digital, viral, Internet.

ABSTRACT

Advertising is not a modern phenomenon. Its origins date from antiquity when the first public criers appeared advertising products. Nevertheless, with the pass of time, advertising developed till reaching the digital era, where we find advertisements and viral advertising. In this study we analyse how the first advertisements were introduced on the net, and how they evolved till nowadays with the help of *Facebook* and *YouTube*. We also offer a classification of advertisements and the media where they may appear. Digital advertising is a communicative process which uses linguistic strategies and as such we cannot forget the influence it has and receives from written conventional advertising.

Keywords: Advertising, digital, viral, Internet.

Nuevas tecnologías y escritura académica¹

LAURA LLANOS CASADO
Universidad de León
lllac@unileon.es

MILKA VILLAYANDRE LLAMAZARES
Universidad de León
mwill@unileon.es

1. INTRODUCCIÓN

El proyecto de apoyo a la innovación docente “Escritura académica” (PAID, 2013), dentro de la línea de actuación “diseño o renovación de metodologías docentes activas, que fomenten la participación del estudiante en el aprendizaje en las titulaciones oficiales de grado y máster de la ULE”, surge para intentar dar respuesta a los problemas detectados en los trabajos de estudiantes de grado dentro del Espacio Europeo de Educación Superior (EEES). En el origen de estos está, por una parte, el desconocimiento de la norma académica y, por otra parte, la falta de referentes en lo que a la redacción de trabajos escritos se refiere. Así, son frecuentes los errores lingüísticos en todos los niveles: faltas de ortografía, discordancias, mal uso de pronombres y tiempos verbales, escaso empleo de conectores textuales, falta de coherencia entre las partes del discurso, inadecuación léxica, ausencia de referencias bibliográficas o forma incorrecta de citarlas, etc. Esta situación repercute de manera negativa en los escritos académicos que los estudiantes presentan a lo largo de su carrera: trabajos, exámenes, memorias de fin de grado, etcétera.

En relación con la situación descrita y teniendo en cuenta que el trabajo de fin de grado (TFG) es una materia común en todas las ramas del conocimiento, se ha considerado necesaria una propuesta que, sirviéndose de las tecnologías de la información y la comunicación así como de recursos gratuitos –en consonancia con las nuevas tendencias metodológicas de enseñanza– permita a los estudiantes perfeccionar su expresión escrita en los múltiples trabajos que tienen que redactar y presentar a lo largo de la carrera.

¹ Este trabajo se enmarca dentro del Proyecto “Escritura académica” (PAID 2013) de la Universidad de León. Además de las autoras, han colaborado en el mismo: Salvador Gutiérrez Ordóñez, Marina Maquieira Rodríguez, Carmen Lanero Rodríguez, Noelia González Verdejo, Marta Alhama Allende y Antonio Barreñada García.

La parte fundamental de la investigación se ha centrado en la revisión de las estrategias fundamentales (planificación, escritura, revisión) y los mecanismos lingüísticos (reglas de puntuación, uso de conectores, mecanismos de objetivación, etc.) necesarios para abordar con éxito la redacción de textos propios del ámbito académico, teniendo siempre en mente como objetivo final el TFG.

De forma esquemática, los objetivos específicos a los que se ha atendido durante el desarrollo del proyecto se resumen en los siguientes puntos:

- Valorar la importancia de la redacción.
- Mejorar la producción escrita en general.
- Reducir el número de errores lingüísticos.
- Conocer las normas que rigen la expresión escrita en español.
- Fomentar el uso de recursos bibliográficos.
- Fomentar el uso de otros recursos relacionados con las nuevas tecnologías (diccionarios electrónicos, blogs, páginas web, herramientas de trabajo colaborativo, portafolio electrónico, Moodle institucional...).
- Redactar textos con adecuación ortotipográfica (grafía, acentuación, puntuación), gramatical (morfológica y sintáctica), léxica y discursiva.
- Fomentar el trabajo autónomo del estudiante.
- Estimular el uso de nuevas tecnologías por parte del profesorado de la Universidad de León (ULE).

2. FASES DEL PROYECTO

La primera fase del proyecto se ha desarrollado durante el segundo semestre del curso académico 2012-2013 con alumnos de la asignatura "Lengua española y medios de comunicación" de tercer curso del Grado en Educación Social. El grupo estaba constituido por ocho hablantes nativos de español y diez sinohablantes. La secuenciación llevada a cabo ha sido la siguiente:

1. Como punto de partida, se realizó una evaluación inicial, para la cual se pidió a los estudiantes que redactaran un texto académico manuscrito² –el comentario de un artículo periodístico³– de un tema de interés relacionado con su ámbito de estudio. Para esta tarea no se les dio ninguna indicación específica, ya que precisamente el objetivo era detectar los problemas más generalizados en sus escritos, sin que ellos fueran conscientes del objeto de la investigación.
2. A continuación, en una primera corrección, se marcaron y aislaron las incorrecciones más frecuentes, clasificándolas según su tipología (con etiquetas del tipo: "discordancia", "léxico", "repetición", "ausencia de estructuradores del discurso", etc.), como se observa en la figura 1. Tras el análisis, sistematización y clasificación de los principales tipos de errores, se observó que una gran parte de ellos giraba en torno a tres aspectos:

² El motivo de preferir en esta fase textos manuscritos es que, de esta manera, se evita la posible influencia de los correctores automáticos incorporados en los procesadores de texto.

³ "El que escriba 'habrir' no debería graduarse" (de Elisa Silió, *El País*, 16/02/2013).

“El que escriba ‘habrir’ no debería graduarse”

Análisis de texto:

Nos encontramos ante un artículo publicado en el diario “El País” el 16 de Febrero del 2013 por una colaboradora habitual de dicho diario, Elisa Silló.

La intención de la autora es hacernos reflexionar este grave problema, el incorrecto uso del castellano por parte de la población joven. Para ello, hace uso de innumerables ejemplos de relatos de expertos académicos y de datos científicos (Barómetro de hábitos de lectura 2012 y Estudio Internacional del Progreso en Comprensión Lectora).

Halla también que el problema no es únicamente ortográfico, si no también sintáctico. Hace hincapié en que el análisis y los acentos importan tanto como las faltas.

La solución que propone el grupo de los docentes a la mala ortografía es que tienen que bajarse la nota.

Comentario [M1]: Mal uso de mayúsculas

Comentario [M2]: Repetición léxica

Comentario [M3]: Falta de preposición

Comentario [M4]: Falta de ortografía

Comentario [M5]: Discordancia

Fig. 1. Texto con marcas de corrección.

- Cuestiones ortotipográficas motivadas, bien por el desconocimiento de las normas y convenciones vigentes, bien por descuidos (estos sin relevancia para el propósito de nuestro estudio).
 - Cuestiones morfosintácticas relacionadas, por una parte, con el mal uso de los pronombres átonos pronominales (casos de laísmos, leísmos y loísmos, motivadas por dudas sobre el régimen verbal o por la procedencia geográfica de los hablantes) y, por otra parte, con la presencia de discordancias nominales y verbales. En cuanto a estas últimas, se concluyó que se debían, en la mayoría de los casos, a la amplia distancia entre el referente y el elemento con el que aquel debía concordar o a la preeminencia de la concordancia *ad sensum* sobre la gramatical.
 - Cuestiones relacionadas con la estructuración del discurso. Aquí, los problemas se centraban en la excesiva longitud de los párrafos –además, mal puntuados– y a la ausencia de marcadores textuales.
3. Después, se devolvieron los escritos a los alumnos para su revisión individual, previa a la entrega de la versión definitiva. Fue en este momento cuando se procedió a presentar el proyecto a los estudiantes, pidiéndoles su colaboración y destacando la importancia de una buena redacción. De este modo fueron conscientes de que cometían más errores de los que inicialmente pensaban, errores que, además, coincidían en la mayoría de los casos.
 4. En una posterior sesión presencial en grupo, se llevó al aula una selección de muestras de sus propios escritos para que ellos mismos propusieran redacciones alternativas a las oraciones o párrafos que contenían

incorrecciones de algún tipo. Solo después de su intervención se explicaba el porqué del error.

5. Con el objetivo de intentar reducir estas incorrecciones, las clases se estructuraron precisamente en relación con las áreas problemáticas: la ortografía, la morfosintaxis y el nivel discursivo. Así, para cada una de ellas, se elaboraron fichas y presentaciones con explicaciones teóricas, se propusieron ejercicios interactivos de corrección automática y se proporcionaron recursos bibliográficos y otros materiales de referencia disponibles de forma gratuita en Internet (como la página de la Real Academia Española: *DRAE*, *DPD*, *CREA*; la de *Fundéu*; el diccionario *Clave*; *Lexipedia*, etcétera).
6. Por último, para valorar la incidencia de las actividades llevadas a cabo durante el semestre, se propuso la redacción de un texto en una sesión presencial. Los resultados pusieron de manifiesto que, si se llama la atención de los estudiantes sobre la existencia de errores en sus propias redacciones y estos se convierten en tema de reflexión explícita en el aula, es posible producir escritos de mayor calidad.

En resumen, de forma esquemática, las fases en que se ha desarrollado el proyecto han sido las siguientes:

1. Evaluación inicial: primera versión.
2. Clasificación y tipología de errores.
3. Revisión individual de la primera versión (no presencial).
4. Revisión presencial en grupo.
5. Elaboración de materiales de apoyo:
 - Cuestionarios.
 - Fichas teóricas.
 - Recursos en Internet.
6. Evaluación final.

3. DESARROLLO DEL TRABAJO

Como ya se ha señalado, el proyecto se inició con la lectura de un texto periodístico que, precisamente, trataba el tema de la preocupante presencia de errores ortográficos entre estudiantes universitarios (Silió, 2013). A raíz de los comentarios sobre este artículo, se constató que las incorrecciones más frecuentes se pueden sistematizar en la tipología que se ofrece a continuación:

a) Errores ortotipográficos.

1. Uso inadecuado de la coma (separando sujeto y verbo, o verbo y complementos).
 - *El profesores López, requirió a sus alumnos que trajeran...*
 - *La autora, recurre a encuestas oficiales y datos reales.*
 - *Es importante destacar que muchos profesores de ciencias, no le dan demasiada importancia a las faltas de ortografía.*
 - *La autora quiere mostrar al lector, que cada vez es más frecuente en los estudiantes universitarios cometer faltas.*

2. Errores de acentuación.

- *...dependiendo de a que se refieran...*
- *Uno de los últimos párrafos trata de los medios de comunicación y como influyen en la ortografía.*
- *Tanto el menos (signo matemático) como el guión (encargado de unir dos palabras, entre otros usos) no se encuentran presentes en el texto.*

b) Errores morfosintácticos.

3. Discordancias nominales.

- *Reproducción textuales*
- *Tecnicismo propios*
- *Uno de los opiniones se menciona...*
- *Se citan opiniones de los profesores, quien reclaman fuertemente que no...*
- *Empleo de tecnicismo propios de las materias que tratan...*
- *En la primera párrafo se conduce el tema...*
- *Como se cita en el texto, probablemente la falta de escribir "abrir" con "h" esté debido a que muchos...*
- *También destacan las comillas sencillas en palabras o citas que se encuentran dentro de una frase entrecomillada y que necesita ser destacada...*
- *(...) las noticias que contiene son de índole muy diverso (...)*

4. Discordancias verbales.

- *La solución que propone el grupo de los docentes a la mala ortografía es que tienen que bajarse la nota.*
- *El título y el subtítulo siempre se concentran en el tema y organiza el artículo bien.*
- *Los medios digitales condiciona la forma de escribir de los jóvenes.*
- *Usa palabras derivadas más largas cuando existe las palabras simples.*
- *Además, en los subtítulos se enumera opiniones sobre este problema.*
- *Sin embargo, todos los periódicos sirven sus intenciones, sea buenas o malas.*
- *...entre las características morfosintácticas se pueden destacar el uso de oraciones interrogativas.*

5. Falta de preposiciones o uso erróneo de estas.

- *Una foto de jóvenes jugando sus móviles.*
- *Transmite tendencia de rapidez.*
- *Los medios digitales sí tienen influencia a la forma de escribir de los jóvenes.*
- *Los jóvenes deben escribir en papel para reducir el impacto de los productos electrónicos para sus formas de escribir.*
- *La gente piensa más del segundo significado.*

6. Indicativo por subjuntivo.
 - *Se colocan imágenes y diagramas para que nos da una sensación de objetividad.*
 - *En el texto se usa la misma caligrafía para que se mantiene la identidad.*
 7. Discordancias en el uso de los pronombres.
 - *Propone la idea del escritor y, después, el artículo comienza a explicarlo.*
 8. Uso inadecuado de la conjunción *ni* cuando el primer elemento no está negado.
 - *Por eso omiten el punto final ni escriben con mayúsculas.*
 9. Uso inadecuado del adverbio *también* precediendo a la conjunción *sino* (y confusión con la secuencia formada por la conjunción *si* y el adverbio *no*).
 - *El título también refleja los sentimientos del autor, si no que también refleja la esencia del artículo.*
 10. Uso del sufijo *-mente* cuando aparecen dos o más adverbios coordinados.
 - *Ayudan al lector a entender el artículo rápidamente y precisamente.*
 - *Para que la gente lo pueda entender claramente y fácilmente.*
 - *Aparecen en los periódicos muchos usos erróneos del lenguaje, inadvertidamente o deliberadamente.*
 11. Uso del indeterminado *un* ante el adjetivo *otro*.
 - *El uso de neologismo es una otra característica del periódico.*
 12. Confusión entre los verbos *ser* ~ *estar*.
 - *El hábito de la lectura en España también es relacionado con el tema.*
 13. Regularización de sufijos irregulares.
 - *Admitible por admisible.*
 14. Uso del infinitivo fático.
 - *Una vez observado el último cuadro (...), decir que, en mi opinión, el problema radica en el propio sistema educativo.*
 - *Por si esto no fuera suficiente gancho para el lector, recordar que, como se ha expuesto, en el primer párrafo del texto la autora utiliza (...).*
- c) Errores discursivos.
15. Repeticiones y pobreza léxica.
 - *Se usan diferentes fuentes en diferentes sitios del artículo.*
 - *El título de un buen artículo debe ser capaz de atraer la atención del lector, lo que permite a los lectores leer su artículo con el interés de seguir leyendo el artículo. En la última parte del artículo (...).*
 - *La opinión de la mayoría de los profesores que opinan en este texto.*

- *...cosa que también llama la atención ya que precisamente "habrir" no sería una falta producto de los SMS ya que en los mensajes cortos lo que se pretende es la brevedad.*

Hay que puntualizar que la distribución de estos errores no es la misma entre los hablantes nativos y los no nativos. Característicos de estos últimos son: las discordancias nominales (en los sinohablantes, por la falta de flexión en su lengua), la ausencia o uso incorrecto de preposiciones, la confusión de formas verbales (indicativo/subjuntivo, tiempos verbales, *ser/estar...*), la combinación de *un* con *otro* (*un otro*), la coordinación de adverbios acabados en *-mente* sin eliminar este sufijo en uno de ellos *y*, en el caso de los adjetivos deverbales, el empleo de una base léxica que no se corresponde con la del derivado en español. Sin embargo, cuando los hablantes son nativos, las incorrecciones afectan, sobre todo, al plano discursivo; además, son especialmente frecuentes las discordancias verbales, la mayoría de ellas debidas a la presencia de incisos entre los términos correferentes o a la gran distancia entre estos. En este sentido, llama la atención el hecho de que los hablantes nativos apenas utilizan mecanismos de cohesión, mientras que los no nativos estructuran mejor sus discursos, quizá porque han recibido una formación explícita sobre el uso de marcadores y otros recursos textuales. Otros problemas observados se refieren a calcos semánticos –en especial del inglés (*soportar* por *apoyar*, como en “*Se citan las opiniones de alguien profesional para soportar la idea principal*”)– y al uso de archisílabos⁴ (*culpabilidad* por *culpa*, como en “*atribución de la culpabilidad a las redes sociales*”), ambos ampliamente documentados en los medios de comunicación, cuya influencia sobre la lengua se deja sentir cada vez más. Asimismo, la falta del dominio léxico se traduce en frecuentes usos inadecuados de términos por desconocimiento de su significado (como *abocar* en “*Las opiniones del profesorado sobre el tema son todas negativas, abocando toda la responsabilidad a los alumnos*”).

Para abordar de forma específica algunos de estos temas, se procedió a la elaboración de una serie de cuestionarios de respuesta múltiple, con una doble finalidad: por una parte, medir la cantidad y cualidad de los errores de los estudiantes; y, por otra parte, servir al profesor como guía para las explicaciones teóricas y el diseño de ejercicios prácticos de refuerzo.

Estos materiales se pusieron a disposición de los alumnos en el Moodle institucional de la asignatura. No obstante, a fin de hacerlos accesibles a un público más amplio, se han volcado también en el sistema de autor Hot Potatoes, y el resultado está disponible en el blog creado a tal efecto⁵. Como ejemplo, la figura 2 refleja un test de ortografía que incluye retroalimentación: muestra a los estudiantes la solución e indica la explicación del fenómeno, tomando como referencia las distintas ediciones de la *Ortografía* académica.

El análisis minucioso de las respuestas de este test nos permitió constatar que muchos de los errores obedecían al desconocimiento de las normas de ortografía básicas y no solo al de las nuevas, como sería de esperar. De hecho, en el diseño de la prueba la mayoría de ítems se había centrado en estas últimas,

⁴ Para un estudio más completo sobre el fenómeno del sesquipedalismo, véase Somoano Rodríguez (2012: 130-172).

⁵ Este cuestionario está accesible, al igual que otros de los elaborados, en el blog de apoyo empleado para el proyecto.

15. La globalización es ante todo un fenómeno antropológico: la humanidad se descubre _____ especie.

Seleccione una o más de una:

- a. como
 b. cómo

Según la *Ortografía básica* (pp. 48-51), *como* se escribe sin tilde cuando funciona a modo de preposición, esto es, cuando introduce un sustantivo o un grupo nominal haciéndolo depender de una palabra anterior: *Está satisfecha de su trabajo como profesora.*

La respuesta correcta es:
 como

16. Son cuatro temas que van a dar _____ hablar.

Seleccione una o más de una:

- a. qué
 b. que

Según la *Ortografía básica* (pág. 51), *que* es átono y se escribe sin tilde cuando forma parte de locuciones o expresiones: *dar que hablar/pensar, hay que ver, ni que decir tiene, etc.*

La respuesta correcta es:
 que

Fig. 2. Cuestionario en línea de ortografía.

mientras que las cuestiones referidas a las primeras ocupaban un lugar marginal. Una muestra significativa de esta tendencia lo ilustra la siguiente figura (Fig. 3), en la que se compara el número de respuestas correctas a propósito de la ortografía de *hui*⁶ y de *paraguas*⁷.

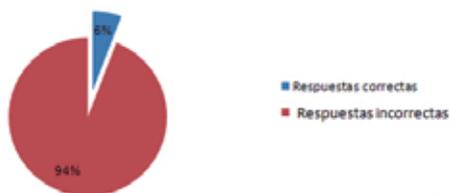
Los cuestionarios se completaron con fichas teóricas para desarrollar en profundidad los aspectos más problemáticos a los que se enfrentan los estudiantes. La que se muestra a continuación (Fig. 4) fue elaborada por el profesor Gutiérrez Ordóñez⁸ para este proyecto.

⁶ Según las nuevas normas ortográficas (RAE y ASALE, 2010), *hui* constituye un caso de dip-tongo ortográfico y, por lo tanto, una palabra monosílaba que, en consecuencia, no se acentúa.

⁷ En este caso, la mitad de los estudiantes consideró que la opción correcta era la que incluía una diéresis (**paragüas*).

⁸ Agradecemos especialmente la participación del profesor Gutiérrez Ordóñez, responsable de la coordinación de la *Ortografía* de la Real Academia Española, cuyas indicaciones han sido de gran ayuda para el desarrollo de esta propuesta de innovación docente.

Hui: desconocimiento de las nuevas normas ortográficas



Paraguas: normas ortográficas básicas



Fig. 3. Gráfico ilustrativo de la ocurrencia de errores.

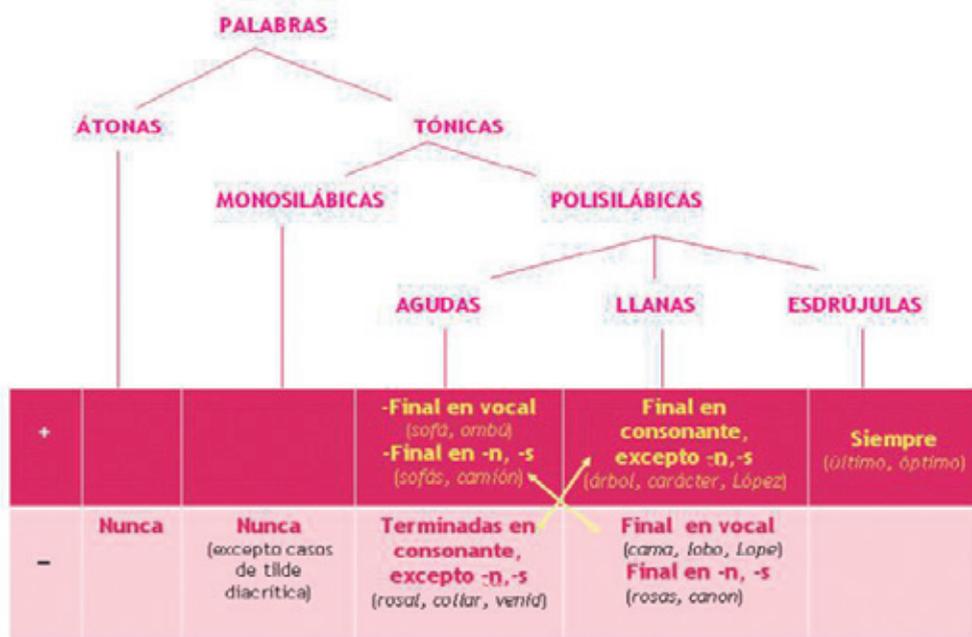


Fig. 4. Ficha teórica elaborada por Salvador Gutiérrez Ordóñez.

Puesto que el objetivo último es la mejora de la escritura en el contexto académico, las herramientas anteriores se aplicaron a lo largo de todo el semestre para poder constatar la evolución de los alumnos. Así, se les pidió que elaboraran un texto, cuya redacción pasó por diferentes fases que se enumeran a continuación.

- Entrega de una primera versión: aquí se observaron los errores destacados en la tipología (vid. supra).
- Primera versión corregida: los errores presentes en los textos se marcaron según un código que reflejaba el tipo de incorrección.

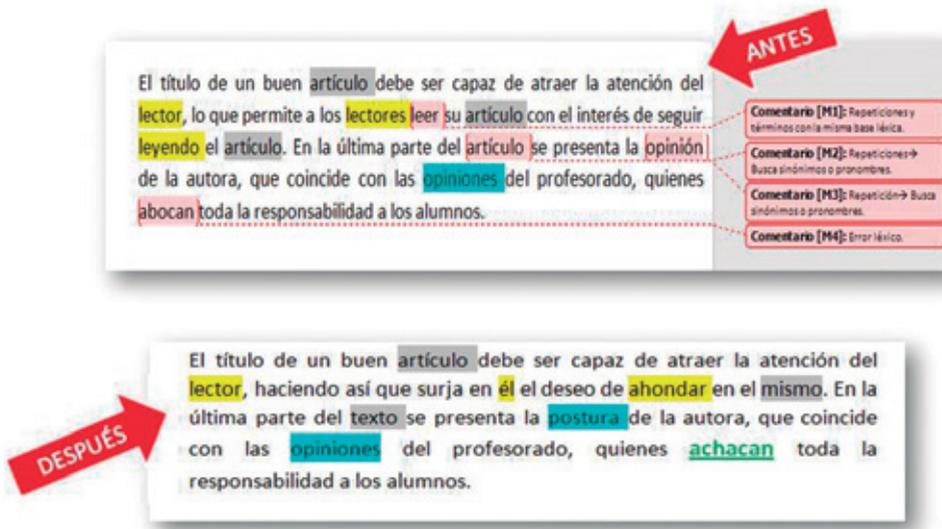


Fig. 5. Imagen comparativa de las fases inicial y final del texto.

- Entrega de la versión final: en esta, se habían revisado los fragmentos señalados en la primera versión, y se había optado por una nueva redacción.
- Nueva y última corrección: en este momento, el profesor procede a la corrección y evaluación de los textos que, aunque todavía contenían algunos errores, habían mejorado respecto a la propuesta inicial.

En la figura 5 se puede comparar la diferencia entre la primera versión del texto entregada por un alumno y la versión final, tras haber pasado por las fases anteriormente descritas.

4. CONCLUSIONES PROVISIONALES

Como primeros resultados del proyecto, se pueden destacar algunos puntos:

1. Los errores que más fácilmente se subsanan son los del nivel ortotipo-gráfico. Por una parte, en muchos casos se trata de normas ya conocidas pero no suficientemente asimiladas; por otra parte, son mucho más fáciles de identificar y corregir. No obstante, conseguir una buena puntuación ha sido el aspecto más difícil, ya que, como apunta Figueras (en Montolío, 2000: 77), "no resulta posible establecer reglas de cumplimiento obligado para todos los casos".
2. Dentro del plano morfosintáctico, las mejoras más notables se aprecian en el área en la que la incidencia de errores es más alta, la de la discordancia verbal. Tan pronto como el estudiante entiende dónde está el origen de este tipo de incorrección, es capaz de evitarla.
3. El escaso empleo de marcadores y otros mecanismos de coherencia y cohesión obedece a la poca atención otorgada al nivel textual en ciclos educativos anteriores. Sin embargo, una vez que se puso de relieve su

importancia y se abordó su tratamiento en el aula, los propios estudiantes tomaron conciencia de la necesidad de su empleo para expresar con mayor claridad el hilo argumental de sus discursos.

4. Otro tipo de errores, como las repeticiones léxicas, se deben a la falta de revisión del alumno de sus propios escritos. En cuanto se les señala, de forma visual, el abuso de un mismo término o de palabras con la misma raíz, son capaces de corregir el texto usando sinónimos o pronombres.

En definitiva, esta forma de trabajar permite realizar un seguimiento personal del alumno⁹, al tiempo que potencia el aprendizaje autónomo y colaborativo. El trabajo se desarrolla de forma escalonada, de tal manera que los estudiantes pueden ir incorporando, progresivamente, las sugerencias que los profesores, mediante tutorías individuales o en grupo, les indican. El objetivo no es tanto que el docente corrija los errores sin más, como que sea el propio alumno quien tome conciencia de los mismos y revise sus redacciones. De este modo, se implica más en el proceso de aprendizaje, percibe la progresiva mejora de su producción escrita, con la motivación que ello conlleva, y desarrolla estrategias que puede emplear en el futuro para mejorar sus textos.

5. TRABAJO FUTURO

Dados los buenos resultados obtenidos hasta el momento, está previsto continuar con el proyecto durante el primer semestre del curso académico 2013-2014; en este caso, aplicando el plan de trabajo descrito a las asignaturas "Español correcto: expresión oral y escrita", común al primer curso de todos los grados de la Facultad de Filosofía y Letras, y "Talleres de lengua y literatura", del tercer curso del grado en Educación Primaria de la Facultad de Educación.

Conscientes de que la buena redacción no se limita ni a las asignaturas más directamente vinculadas con la lengua ni al ámbito de las Humanidades, sería deseable ampliar el ámbito de trabajo y trasladar los resultados:

- a la redacción de las memorias de prácticas, diarios de laboratorio, etcétera;
- a otras asignaturas y a los trabajos de fin de grado;
- a otros estudios;
- a otros niveles educativos;
- a la redacción de trabajos de fin de máster, trabajos de investigación o artículos científicos.

En conclusión, la propia naturaleza de nuestro objetivo (escritura de un trabajo final escrito, con vistas a mejorar la redacción del TFG) hace de esta propuesta una clara candidata para ser incluida en cualquier ámbito de conocimiento. Como es sabido, en la ordenación de las enseñanzas universitarias oficiales se determina que las nuevas titulaciones deben incluir un Trabajo de Fin de Grado orientado a la evaluación de competencias asociadas a la titulación. Además, la evaluación de este TFG no se limita a la elaboración del mismo, sino que su presentación tendrá también repercusión en la nota final del alumno. Por ello, las técnicas de mejora de la expresión oral y las estrategias que pueden orientar

⁹ Trabajar con grupos reducidos ha sido fundamental para llevar a cabo el proyecto.

al alumno en la presentación de trabajos académicos podrán constituir un nuevo enfoque de este mismo proyecto en sucesivas etapas, así como su traslado a otros formatos, como los COMA (Cursos *Online* Masivos y Abiertos).

BIBLIOGRAFÍA

- Aleza Izquierdo, Milagros (coord.), *Normas y usos correctos en el español actual*, Valencia, Tirant Humanidades, 2011.
- Arteta, Aurelio, "Arrecian los archisílabos", *El País*, 10 de agosto de 2005, <http://elpais.com/diario/2005/08/10/opinion/1123624805_850215.html> [28/02/2013].
- Cortés Moreno, Maximiano, "Chino y español: un análisis contrastivo", *Carabela*, 52 (2002), pp. 77-98.
- Gómez Manzano, Pilar *et al.*, *Ejercicios de gramática y de expresión (con nociones teóricas)*, 2ª ed., Madrid, CEURA, 2007.
- Gómez Torrego, Leonardo, *Hablar y escribir correctamente. Gramática normativa de español actual* (2 tomos), Madrid, Arco/Libros, 2006.
- Martínez García, Hortensia, *Construir bien en español. La corrección sintáctica*, Oviedo, Nobel, 2005.
- Moliner, María, *Diccionario de uso del español*, Madrid, Gredos, 1966-1967.
- Montolío Durán, Estrella (coord.), *Manual práctico de escritura académica*, Barcelona, Ariel, 2000.
- Real Academia Española, *Diccionario de la Lengua española*, Madrid, Espasa Calpe, 2001.
- Real Academia Española, *Diccionario panhispánico de dudas*, Madrid, Santillana, 2005.
- Real Academia Española y Asociación de Academias de la Lengua Española, *Manual de la nueva gramática de la lengua española*, Madrid, Espasa, 2010a.
- Real Academia Española y Asociación de Academias de la Lengua Española, *Ortografía de la lengua española*, Madrid, Espasa, 2010b.
- Real Academia Española y Asociación de Academias de la Lengua Española, *Nueva gramática básica de la lengua española*, Madrid, Espasa, 2011.
- Real Academia Española y Asociación de Academias de la Lengua Española, *Ortografía básica de la lengua española*, Madrid, Espasa, 2012.
- Serrano Serrano, Joaquín, *Guía práctica de redacción*, Madrid, Anaya, 2002.
- Silió, Elisa, "El que escriba 'habrir' no debería graduarse", *El País*, 16 de febrero de 2013, <http://sociedad.elpais.com/sociedad/2013/02/16/actualidad/1361037969_843190.html> [28/02/2013].
- Somoano Rodríguez, Julio, *El papel de la radio, la televisión y la prensa en la normalización en la lengua española del eufemismo, el sesquipedalismo y el malapropismo*, Madrid, Universidad Complutense de Madrid, 2012.

RECURSOS EN INTERNET

- El castellano.org*, <<http://www.elcastellano.org>> [04/05/2013].
- En román paladino*, <<http://canales.larioja.com/romanpaladino>> [04/05/2013].
- Escritura académica*, Blogger, <<http://escrituraacademicaule.blogspot.com.es>>.
- Fundación del español urgente, <<http://www.fundeu.es>> [04/05/2013].
- Hot Potatoes, Universidad de Victoria, Canadá, <<http://hotpot.uvic.ca>>.
- Lexipedia, <<http://www.lexipedia.com/>> [04/05/2013].
- Moodle, Universidad de León, <<https://agora.unileon.es/auth/saml/login.php>>.
- Real Academia Española, *Banco de datos (CREA)* [en línea], *Corpus de referencia del español actual*, <<http://www.rae.es>> [04/05/2013].

Real Academia Española, *Diccionario de la lengua española* [en línea], <<http://www.rae.es>> [04/05/2013].

Real Academia Española, *Diccionario panhispánico de dudas* [en línea], <<http://www.rae.es>> [04/05/2013].

SM, *Diccionario Clave*, <<http://clave.smdiccionarios.com/app.php>> [04/05/2013].



RESUMEN

La producción escrita en español presenta cada vez más problemas derivados, por una parte, del desconocimiento de la norma académica y, por otra, de la falta de referentes en lo que a la redacción de trabajos se refiere. Son frecuentes los errores lingüísticos en todos los niveles. Esta situación repercute de forma negativa en los escritos de los estudiantes universitarios, tanto si se trata de nativos como de no nativos.

En relación con esta situación, la presente contribución avanza los resultados provisionales del proyecto de innovación docente, desarrollado en la Universidad de León para la mejora de la escritura académica, con especial atención a la incidencia de las nuevas tecnologías y recursos en línea.

Se describen, asimismo, las distintas fases del trabajo llevadas a cabo y cómo este se ha traducido en una reducción del número de errores lingüísticos, mejor conocimiento de las normas que rigen la expresión escrita en español y uso adecuado de las convenciones ortográficas, gramaticales, léxicas y tipográficas.

Las conclusiones nos inducen a afirmar que un análisis crítico basado en muestras y casos reales permite diseñar materiales apropiados. Por otro lado, se fomenta la autonomía del estudiante y se estimula el uso de nuevas tecnologías, tanto en la práctica docente como por parte del alumnado.

Palabras clave: Nuevas tecnologías, escritura académica, recursos en línea, innovación docente, tipología de errores.

ABSTRACT

Writing is becoming more and more a problematic issue for university students. This situation is due, on the one hand, to the fact that they fail to know some grammar, spelling and other typographical conventions and, on the other hand, to the lack of references on text composition.

Mistakes are frequent in every linguistic level. As a result, the writings, no matter whether produced by native or non-native speakers, show a decrease in their quality and efficiency.

In this regard, the present contribution aims at describing the provisional results obtained from the teaching innovation project –developed at the Spanish Department of the University of León, to improve academic essays–, focusing on new technologies and online resources as an aid for writing.

The different phases taken are also commented as well as the improvement observed. The most evident consequences are the reduction in the number of linguistic mistakes, a better knowledge of text organisation in Spanish, an appropriate use of orthographical and grammatical rules, and the selection of lexical entries.

So far the conclusions allow us to state that a thorough analysis based on the use of real data is crucial to design adequate teaching materials. In addition, new technologies help to encourage both learner autonomy and teacher use of innovative tools.

Keywords: Information technology, academic writing, online resources, teaching innovation, error typology.

De cómo Ramón Gómez de la Serna se convirtió en un fan de Twitter: un ejemplo de textualidad digital

ÁLVARO LLOSA SANZ
Hobart and William Smith Colleges
llosasanz@hws.edu

A través de las próximas páginas, bajo el título “De cómo Ramón Gómez de la Serna se convirtió en un fan de Twitter: un ejemplo de textualidad digital” voy a contarles la historia de cómo un libro impreso puede devenir en algo diferente a su formato tradicional al mismo tiempo que respetamos su objetivo original: diseminar y compartir la ficción con su audiencia.

Como bien sabrá el lector, Ramón Gómez de la Serna fue un escritor español vanguardista, dramaturgo, y agitador de conciencias. Cuando en 1955 publicó su *Total de greguerías*, una recopilación de unas 10.000 micropiezas poético-humorísticas, no había internet ni redes sociales virtuales. Sin embargo, hoy voy a explicarles cómo Ramón, que murió en el año 1963, se convirtió en un fan de twitter en el siglo XXI, y cómo sus greguerías abandonaron un artilugio tecnológico llamado libro para abrirse a una nueva vida gracias a cierto proceso de transformación que ha hecho pervivir y difundir las obras de arte durante siglos: la textualidad.

“Conviene asimismo tener en cuenta que la lectura es siempre una práctica encarnada en ciertos gestos, espacios y hábitos” (Cavallo y Chartier, 1998: 12). Las obras de ficción no son inmutables, pero perduran precisamente porque sufren cambios de lugar, no sólo a causa de las versiones realizadas por diversos artistas en diversos medios, sino también porque toda obra de ficción es diferente para cada lector en una circunstancia de espacio y tiempo particulares, en el momento de su lectura y en cómo y dónde la ejecuta. El texto mismo, al ser difundido en un soporte o medio específico que se dirige a un público concreto, la edición misma es parte de ese evento de lectura, de la actividad y la actuación lectora. Hay por tanto un acto particular de autoridad editorial al elegir tal o cual medio o formato de difusión. La textualidad se fundamenta en ese carácter de cambio de acuerdo a un espacio performativo del texto: “An edition, like a text, is merely a site of passage of a work of literature; a site in which instabilities are both made and made manifest” (Grigely, 1995: 118).

Como soportes de memoria, tanto los libros como los espacios digitales contienen en su construcción una retórica específica y particular adecuada a una exploración que nos ofrezca una experiencia lectora con sentido. Si bien la pá-

gina, los párrafos, los índices y las notas al margen regulan el mundo del libro desde su desarrollo medieval (Cavallo y Chartier, 1998: 53), respecto a los espacios digitales, debemos destacar que el principio de interactividad sobre la pantalla regula nuestra relación con la máquina y se diversifica en tres niveles de interacción: primero, la navegación hipertextual; segundo, la exploración inmersiva; tercero, la interacción registracional (Lister, 2009: 22). Estos tipos de interacción son dominantes pero no exclusivos de los soportes digitales, puesto que pueden hallarse en soportes analógicos también, como veremos ahora con algunos ejemplos. Sólo así podría hallarse un hilo conductor entre la retórica impresa y la digital, si entendemos el texto, independientemente del soporte que lo alberga, como cibertexto (Aarseth, 2004: 118-121), es decir, como máquina combinatoria donde el contenido es un engranaje de interacciones entre sus materiales y su puesta en marcha por un lector usuario.

En primer lugar, la navegación hipertextual nos permite elegir qué fragmento de información nos gustaría leer. Nos ofrece esta posibilidad porque cada pieza informativa se ha integrado en una base de datos. Un buen ejemplo cotidiano es un periódico en línea, que está conformado por bloques informativos enlazados que nos permiten seleccionar o buscar aquello que queremos leer a continuación. De un modo similar, el índice de contenidos, las notas a pie de página o el índice final en un libro impreso generan una navegación hipertextual. En definitiva, leer hipertextualmente es hacer elecciones de lectura en una base de datos estructurada (“to make reading choices in a database”) (Lister, 2009: 22).

En segundo lugar, explorar inmersivamente es la manera en que leemos o identificamos sentidos al recorrer un espacio visual. Un ejemplo de esto es la representación gráfica de la información en un video juego, en el que la información está enlazada con una exploración sensorial del espacio. En breve, consiste en “explore and navigate in visually represented screen spaces” (Lister, 2009: 22). Esta inmersión implica habitualmente la creación de un texto multimodal, en el que imagen, sonido y palabra escrita se unen para narrar con gran efectividad. Los antiguos mapas y los manuscritos iluminados en pergamino y papel fueron medios predigitales basados en la exploración inmersiva. Hoy me centraré más en el aspecto multimodal de esta navegación exploratoria.

Y en tercer lugar, la interacción registracional nos permite añadir nueva información al texto, y nos ofrece “opportunities that new media texts afford their users to ‘write back into’ the text; that is to say, to add to the text by registering their own messages” (Lister, 2009: 22). Un ejemplo fascinante en nuestra cultura digital es la interminable oportunidad de dejar comentarios y reflexiones personales en bitácoras en línea y redes sociales. De igual modo, encontramos este tipo de interacción en los *marginalia* de muchos libros manuscritos e impresos. Por comodidad y extensión de uso, prefiero llamar a este tipo de interacción, interacción social.

En este contexto de interacciones, ¿cómo es posible transformar una edición impresa de las greguerías de Gómez de la Serna en una edición hipertextual, multimodal y social? ¿Y cómo puede actuar adecuadamente la textualidad en el medio digital?

Vamos a explorar, por tanto, esa transformación digital o tercera mediamorfosis (Fidler 54 y 71-80) a través de un proyecto existente en desarrollo. De abril a septiembre de 2011 creé un experimento en línea que se convertiría en una experiencia de edición digital de las greguerías ramonianas. En un principio

pensé que una colección de textos ficcionales tan breves generaría fácilmente una base de datos literaria para una audiencia de lectores en línea. Cada greguería se convertiría en una unidad de lectura, y cada unidad podría ser recuperable por el lector gracias a una caja de búsqueda tras una etiquetación del contenido. De este modo tenía ya el principio hipertextual de la edición, pero pronto me di cuenta que un total de 10.000 greguerías es un corpus enorme como para esperar a transcribirlas todas sin empezar a publicarlas antes. Para solucionar la necesidad de publicar parte del material mientras se iba añadiendo otro, recurrí al concepto de evento vivo de lectura encarnado en la idea de los microblogs.

Por lo tanto, decidí crear inicialmente un blog que contuviera mi base de datos para difundir adecuadamente una greguería cada día de la semana. Aunque al comienzo usé la plataforma Blogger, actualmente puede encontrarse en la siguiente dirección de la plataforma Wordpress, mejor estructurada para combinar categorías y etiquetas: <<http://totaldegreguerias.wordpress.com>>. Cada unidad o greguería se asociaba a palabras clave para poder recuperar después cualquier greguería, usando como ayuda una caja de búsqueda de texto y una nube de etiquetas. Estas palabras clave las constituyen los nombres, acciones verbales, características adjetivales y circunstancias adverbiales contenidas en cada una de las unidades, unificadas de tal modo que si el término *cisne/s* aparece en varias greguerías, sea posible recuperar ambas al mismo tiempo. Desde luego, aquí se me plantea la cuestión de si debería haber áreas temáticas de etiquetas para describir los temas de las greguerías, u otras posibles categorizaciones, pero el proyecto es lo suficientemente flexible y actualizable en este sentido como para poder incorporar nuevas categorías en cualquier momento. Y quizás la apertura hacia un modelo de colaboración masiva y abierta (el *crowdsourcing* defendido por Clay Shirky) pudiera dar resultados sorprendentes y rápidos, y sería muy útil dada la extensión del corpus. En cualquier caso, la edición inicial en proceso cumple al menos con los requisitos esenciales de etiquetación del contenido para generar una red semántica bien delimitada en cuanto a los elementos que se describen en las greguerías. De esta manera, comencé a desarrollar y publicar una edición hipertextual *gota a gota*, recibida por el lector de una manera temporal, diaria en este caso. Nadie va a leer 10.000 greguerías en una semana, pero recibir en el correo una al día es un placer. Y si se quiere más, una visita al blog con su caja de búsqueda y su nube de etiquetas nos permite extraer una greguería a través de las palabras y áreas semánticas disponibles.

Ahora bien, ¿qué hay de la multimodalidad? Si bien el lenguaje es fundamentalmente imágenes (físicas o conceptuales) elaboradas con palabras, en este caso hay que destacar cómo la edición impresa incluye unas 300 greguerías visuales dibujadas por el autor mismo. Sin duda, este aspecto multimodal que ha sido obviado en la mayoría de las antologías, exigía crear un nuevo repositorio exclusivamente para ellas por medio de un sitio en línea para compartir imágenes, y permitir por tanto una exploración visual del universo ficcional de las greguerías. El uso de un repositorio profesional como Flickr (<www.flickr.com>) ofrecía una buena oportunidad que luego demostró poder ampliarse a otro aspecto importante de esta edición piloto, ligado a la participación lectora. Tenía solucionados los problemas de transtextualidad básicos pero me sentía solo. Un libro sin lectores no es un libro.

Para convertir la edición en algo más que una base de datos hipertextual y multimodal, decidí usar las capacidades sociales de diferentes servicios. Pensé que si difundía una unidad al día en una red social, podría generarse un club de

lectura, una comunidad ramoniana de greguerías. Twitter (<www.twitter.com>) es perfecto para este objetivo, ya que es posible seguir a Ramón y su actividad. En este sentido, decidí que Ramón debería contar con su propio perfil, extraído de sus propias palabras en algunos extractos de su autobiografía, para que sus lectores interactuaran directamente con él. Además, creé una línea de tiempo de Twitter que recoge greguerías y otras referencias y Ramón en vivo y en directo hechas por otros usuarios de Twitter. De hecho, sin acudir a la edición específica de la que estamos hablando, Ramón y sus greguerías son ya increíblemente famosos en la red, cada 2 ó 3 horas de media alguien se refiere a él en Twitter y por tanto existe una edición colectiva, inconsciente pero constante, de su obra y pensamiento. En este momento, descubrí que Ramón, con los varios perfiles que ya tiene en la red, se había convertido también en un fan de Twitter.

Me gustaría compartir algunas de las interacciones de los lectores con Ramón en Twitter, derivados de la línea de tiempo que se genera automáticamente recogiendo referencias que lo citan, y en la que yo como editor de Ramón, intervenía de vez en cuando para entablar una relación con los lectores y atraerlos al nuevo perfil y al blog de greguerías.

Retwitteado por R. Gómez de la Serna



Antonio Salazar @mequitoelcraneo · 5 de abr. de 2011
 RT @Camilo_de_Ory Ramón Gómez de la Serna inventó el #Twitter
<http://bit.ly/fGZ84p> Interesante artículo de granadahoy.com

Abrir Responder Retwittear Favorito Más

Retwitteado por R. Gómez de la Serna



Gonzalo de Santiago @Gonzalo14H · 7 de abr. de 2011
 Si Ramón Gómez de la Serna hubiera vivido en la era del Twitter...

Abrir Responder Retwittear Favorito Más

Retwitteado por R. Gómez de la Serna



Javier Sanz @jsanz · 10 de abr. de 2011
 Los mejores tuiteros de la Historia: wp.me/pTYWB-jY vía @victurs Los míos
 Groucho Marx y Ramón Gómez de la Serna

Abrir Responder Retwittear Favorito Más

R. Gómez de la Serna @RGdelaSerna · 9 de jun. de 2011
 ¡Afila lápiz! "Las greguerías son los sacapuntas de las ideas." RT @PacoConut:
 @RGdelaSerna Ay Ramoncete, a ver si me caes mañana en el examen

Abrir Responder Retwittear Favorito Más



Para completar el aspecto participativo en el área multimodal de la edición, creé un grupo en Flickr (<http://www.flickrriver.com/groups/greguerias/pool/>) para que otros usuarios pudieran añadir sus propias greguerías visuales, inspiradas por aquellas que elaboró Ramón. Al ser Flickr un repositorio activo de individuos aficionados a la fotografía, los participantes y sus foto-greguerías son variados y atractivos.

Ahora sí, la edición parecía ser hipertextual, multimodal y social.

Hay una cuestión que quizás preocupa cuando tratamos de nuevas formas de diseminar textos canonizados y adscritos, por tanto, a un medio de difusión muy nuevo y particular. ¿Podemos considerar esta edición como filológica? Desde luego, una edición filológica debe contener un texto preciso y además anotado, que normalmente informa sobre asuntos en torno al texto mismo o sus circunstancias, como en las ediciones anotadas, por ejemplo. Me parece que hay una respuesta evidente para esto cuando hablamos de una edición digital de estas características: a través de la integración de las redes sociales y el blog, no sólo el investigador, sino también otros colaboradores tienen la oportunidad de dejar comentarios informados en cada una de las greguerías. El archivo está vivo y abierto, recuperable en cualquier momento, y por lo tanto en cualquier momento puede rescatarse una de las greguerías y comenzar o seguir una discusión, incluir una anotación, establecer una referencia o dejar una impresión. No obstante, y especialmente para animar a esta tarea, generé una página en Facebook (<https://www.facebook.com/pages/Ram%C3%B3n-G%C3%B3mez-de-la-Serna/136872976385767?sk=wall>) donde crear cómodamente actualizaciones sobre extractos literarios, referencias externas, etc. Fácilmente, extractos y otros materiales que ya están disponibles en línea, o que pueden incorporarse tras recuperarse nuevos documentos sobre el autor en archivos como el de la Universidad de Pittsburgh, se incorporarán en dicha página para su discusión a

través también del blog, donde una categoría diferente a la de las greguerías contiene como base de datos o archivo estas microediciones de materiales reeditados e inéditos. La plataforma es tan versátil y estándar que varios especialistas podrían de hecho incorporarse y trabajar en ella. Sin duda, un grupo colaborador de investigadores apoyados adecuadamente por instituciones es capaz de crear y actualizar, de dinamizar contenidos de calidad e interés para una audiencia global. Debemos tener en mente que vivimos en un universo cultural del entretenimiento y del aprendizaje, donde la convergencia del contenido a través de diversos medios integrados forma parte intrínseca de la esfera de la industria cultural y la actual participación de las audiencias (Jenkins, 2006: 2).

Este tipo de ediciones digitales entendidas como un evento ficcional participativo deberían desarrollarse y estar apoyadas por instituciones o editoriales para garantizar su continuidad, y nos recuerda además dos papeles importantes que justifican la tarea para los investigadores, cuando se enfrentan a la integración digital de soportes, la edición textual y el público. El primero es el del investigador editor como gestor de contenidos: todo aquello conectado con un tema o un autor debería ser difundido a lo largo del tiempo, midiendo bien el momento y la cualidad y cantidad de información, distribuida en bloques breves que tengan sentido de cierta unidad, y trayendo información relevante al momento de publicación. Difundir los primeros párrafos de la *Automoribundia* de Ramón en el aniversario de su cumpleaños puede ser un buen ejemplo, ya que en ellos hace referencia a su momento y lugar de nacimiento. Lógicamente, esto exige atención y conocimiento de la materia literaria, es decir, la dedicación de filólogos profesionales. En segundo lugar, el papel del investigador y editor digital es convertirse en gestor de comunidades virtuales: crear y mantener una audiencia informada y atraída por el contenido que se difunde, porque al fin y al cabo los lectores forman parte ineludible del evento de lectura y de la edición misma. Como dice Ricardo Piglia, “puedo imaginar un mundo sin libros pero no un mundo sin lectores” (2013).

Gardner Campbell dijo en Union College hace unos meses que “la investigación digital es la amplificación de la inteligencia colectiva”. Podríamos alcanzar ese objetivo si repensamos la manera en que difundimos los trabajos de ficción sobre los que investigamos, e incluir otras formas de implicar al público lector en el que se incluyen otros investigadores especializados. En ese sentido, podríamos añadir la existencia potencial de un “excedente cognitivo” (Shirky, 2010) al desarrollo y resultado de nuestras tareas académicas, ya que la audiencia comienza a participar no sólo como lectores pasivos de una ficción, sino también como agentes activos que participan en un evento vivo y colaborativo. Incluso me atrevería a añadir que esto crea un valor cívico compartido con esa comunidad lectora participativa en nuestra tarea como investigadores que abogan por las humanidades digitales.

Por todas estas razones, hoy he querido explicar cómo Ramón Gómez de la Serna se ha convertido irremediablemente en un fan de twitter y ahora confieso mi esperanza de que nos ayude a construir mejores experiencias de lectura digitales. ¿Qué piensan ustedes, creen que podemos intentar con algo más largo, con narrativas más complejas, como las novelas? Para estimular a la reflexión sobre esta última cuestión, dejo aquí el enlace a un video muy popular de 2010 realizado por la empresa Excentric, que nos muestra cómo experimentamos hoy la ficción: “La historia digital de la Navidad” (*The digital story of the Nativity*) narra cómo narrar la historia de la Navidad mediante la integración digital de los

medios actuales. Doce millones de personas vieron el video y están esperando más, mucho más que un texto en el formato pdf para continuar disfrutando de sus ficciones favoritas de todos los tiempos. Como se dice al final del video, puede que los tiempos cambien, pero las emociones son las mismas ("Times change, the feeling remains the same. Excentric: Experience Digital Power").

BIBLIOGRAFÍA

- Aarseth, Espen, "La literatura ergódica", en Domingo Sánchez-Mesa (ed.), *Literatura y cibercultura*, Madrid, Arco-Libros, 2004, pp. 117-145.
- Campbell, Gardner, "What is Digital Scholarship?", NY6 Digital Scholarship Speaker Series, 29 de enero de 2013.
- Cavallo, Guglielmo y Chartier, Roger, "Introducción", en Roberto Bonfill (ed.), *Historia de la lectura en el mundo occidental*, bajo la dirección de Guglielmo Cavallo y Roger Chartier, Madrid-México, Taurus, 1998.
- Excentric, "The Digital Story of the Nativity", *YouTube*, <<http://www.youtube.com/watch?v=GkHNNPM7pJA>> [30/09/2013].
- Fidler, Roger F., *Mediamorphosis: Understanding New Media*, Thousand Oaks, Pine Forge, 1997.
- Gómez de la Serna, Ramón, *Total de Greguerías*, Madrid, Aguilar, 1955.
- Grigely, Joseph, *Art, Theory, and Textual Criticism*, University of Michigan Press, 1995.
- Jenkins, Henry, *Convergence Culture: Where Old and New Media Collide*, New York, New York UP, 2006.
- Lister, Martin, *New Media: A Critical Introduction*, Francis and Taylor, 2009.
- Piglia, Ricardo, Entrevista digital, *El País*, 18 de septiembre de 2013, <<http://www.elpais.com/edigitales/entrevista.html?encuentro=10400>> [30/09/2013].
- Shirky, Clay. "How cognitive surplus will change the world", TED talk, 29 de septiembre de 2010, <<http://www.youtube.com/watch?v=qu7ZpWecIS8>> [30/09/2013].



RESUMEN

Gracias al concepto de textualidad defendido por Joseph Grigely, entendemos la edición de una obra literaria como un evento único que se actualiza en el momento, lugar, circunstancias y público de la edición misma, además de cada acto particular de lectura. Quiero exponer un ejemplo teórico-práctico de edición *online* que se explica bien en el contexto de la actual cultura de la convergencia defendida por Henry Jenkins. Demostraré cómo el *Total de Greguerías* (1955) de Ramón Gómez de la Serna, adaptado al medio digital y sus mecanismos de lectura, explora bien tres tipos de interactividad básicos del medio: la interacción hipertextual, la inmersiva y la registracional. Se crea así una edición de lectura de carácter no lineal, multimodal, y social, además de rigurosamente académica, que reinterpreta el texto literario como cibertexto y lo integra en la cultura participativa de los blogs y las redes sociales. Las tareas necesarias para llevar a cabo con éxito este modelo de edición muestran además que el investigador-editor de literatura cumpliría dos papeles importantes en la elaboración de ediciones digitales accesibles a los lectores: una como gestor de contenidos y otra como gestor de una comunidad lectora. Con este modelo propongo experimentar y repensar la manera en que difundimos la literatura de origen impreso de los autores sobre los que investigamos, incluyendo otras formas participativas de implicar a una audiencia lectora convertida ahora en comunidad lectora digital.

Palabras clave: Textualidad, greguerías, edición digital, prácticas de lectura, interacción.

ABSTRACT

Joseph Grigely coined the term textualterity to explain how an artwork, for example a fiction, could be understood as an event and not merely as an artifact. An artwork changes and is updated depending on the circumstances of creation and dissemination, on the spaces of inscription where is placed, and on the act of reading performed by every reader as a physical individual. This paper shows a practical example of textualterity applied upon a digital online blog-based edition of *Total de Greguerías* (1955), by Spanish avant-garde author Ramón Gómez de la Serna. I explore how textualterity is possible for our digital space when considering three kinds of interactivity: hypertextuality, multimodality, and registrational interaction among users. I show that a digital transformation from the book to the screen is a scholarly practice when we are able to read printed books as a cybertext and every text as a participatory event to share. For being successful to accomplish the task, I conclude that the digital researcher and editor should be a precise content manager and a well prepared community manager as well. In short, this paper rethinks the way we as scholars disseminate printed fiction in our digital era.

Keywords: Textualterity, Greguerías, Digital Publishing, Reading Practices, Interactivity.

Recursos digitales ofrecidos por el SIELAE para el estudio del Siglo de Oro. Creación, gestión y evolución de BIDISO¹

SAGRARIO LÓPEZ POZA
Universidade da Coruña
sagrario@udc.es

ÁNGELES SAAVEDRA PLACES
Universidade da Coruña
asplaces@udc.es

1. INTRODUCCIÓN²

Este trabajo es parte de un bloque de varios integrados en un panel común dedicado a los "Recursos digitales ofrecidos por el SIELAE para el estudio del Siglo de Oro", conseguidos de la colaboración fructífera entre el SIELAE y el Laboratorio de Bases de Datos de la UDC. En primer lugar, Sagrario López Poza presentará al SIELAE (su creación, finalidad y líneas de investigación) y se expondrán dos de esas líneas con más detalle³. A continuación, Ángeles Saavedra Places, profesora del Laboratorio de Bases de Datos de la UDC, presenta ese grupo de investigación, la metodología de trabajo que se ha seguido para la creación y mantenimiento de BIDISO, los aspectos clave en su desarrollo, su evolución, arquitectura, tecnología e infraestructuras actuales.

2. EL SIELAE

La idea de crear el SIELAE (Seminario interdisciplinar para el estudio de la Literatura Áurea Española)⁴ se gestó en 1992 (y se concretó en 1996) al percatarnos de que el estudio de la literatura y la cultura de los siglos XVI y XVII, para una comprensión cabal, precisaba de enfoques no limitados a una disciplina

¹ Este trabajo se inscribe en el proyecto de investigación *Biblioteca digital Siglo de Oro IV* (código FFI2012-3436) financiado por el Ministerio de Economía y Competitividad del Gobierno de España, en el marco del VI Plan Nacional de I + D + i 2008-2011.

² Este trabajo está parcialmente financiado por la Xunta de Galicia (Fondos FEDER), ref. GRC2013/053.

³ Las otras dos líneas (Relaciones de sucesos e Inventarios de bibliotecas) serán tratadas en sendos artículos de esta publicación por Nieves Pena Sueiro y Carlota Fernández Travieso.

⁴ Creado por Sagrario López Poza, catedrática de Literatura Española de la Universidade da Coruña, <<http://www.bidiso.es/sielae/>>.

o área de conocimiento. Había que deshacer las fronteras postizas de competencias de áreas impuestas durante los últimos siglos (y que no existían en el Renacimiento y el Barroco) para acometer de manera más integral el estudio desde perspectivas multidisciplinares.

Además de la certeza de que el enfoque de estudio había de ser multidisciplinar, también teníamos un planteamiento crítico de lo que era una actitud bastante generalizada entre colegas de las áreas de Filología y, más concretamente, de Literatura. Eran muchos los trabajos de crítica literaria basados sencillamente en intuiciones, percepciones, opiniones y posicionamientos del especialista no avalados por una investigación contrastada con datos, y ese tipo de estudios tenían para nosotros poco interés, nos parecía que aportaban poco y que sus conclusiones eran forzosamente percederas.

Sin abandonar los métodos analíticos, críticos o especulativos que caracterizan al área de conocimiento de la Filología Hispánica, queríamos respaldar con datos nuestros estudios y a la vez ayudar a otros en la búsqueda, estudio y difusión de algunos de los campos menos trabajados hasta entonces en la Literatura del Siglo de Oro. Para ello, las nuevas tecnologías fueron la respuesta.

Otro eje central del SIELAE fue desde sus inicios el ánimo de difundir de forma gratuita y abierta todo lo que el grupo había investigado, para que fuera de provecho inmediato para otros.

El grupo ha evolucionado desde comienzos de los 90 (han pasado veinte años desde sus inicios) y más de 50 personas han estado involucradas de un modo u otro en los trabajos que hemos acometido, con 9 proyectos de investigación subvencionados en concurrencia competitiva a nivel estatal.

A lo largo de dos décadas, las líneas de investigación han ido incrementándose, y uno de los logros del grupo es que mantiene, difunde y preserva para el futuro el fruto de sus trabajos, bases de datos o bibliotecas digitales, accesibles a través de Internet con acceso abierto. Esas líneas son:

- La literatura Emblemática española.
- Las Relaciones de sucesos españolas de los siglos XVI-XVIII.
- Los repertorios y misceláneas de erudición humanística (polianteas, florilegios, enciclopedias y diccionarios de lugares comunes...).
- Inventarios y bibliotecas de los Siglos XVI-XVIII.

2.1. Los libros de emblemas⁵

El primer proyecto se focalizó en los libros de emblemas españoles, y fue subvencionado por la Xunta de Galicia. Estas obras encierran un patrimonio cultural

⁵ El primer libro de emblemas, el *Emblematum liber* de Andrea Alciato (1492-1550), que inicia un fértil género editorial y literario, se publicó en Augsburgo, el 28 de febrero de 1531, por el impresor Heinrich Steyner. Era una colección de 104 epigramas en latín, con un título cada uno. Steyner consideró oportuno ilustrar 97 de ellos con grabados. Al parecer encargó los dibujos al pintor Jörg Breu y los tacos xilográficos a Hans Schäußelein. El éxito fue tal, que en abril del mismo año se hacía una segunda edición, y a finales del siglo XVII se habían realizado 171 ediciones de los Emblemas de Alciato, no sólo en latín, sino en francés, alemán, italiano, español y una selección de ellos se tradujo al inglés, de modo que se convirtió en uno de los libros de mayor éxito editorial en la historia de la cultura de Occidente. A imitación de Alciato, se produjeron unos 6.500 libros de emblemas en todas las lenguas europeas.

impresionante, mezclando la palabra y la imagen y no sólo interesan por su contenido literario, político, religioso, botánico, de cultura material y cotidiana, arquitectura efímera, musical, etc. sino que ayudan a comprender mejor aspectos simbólicos de la poesía, la prosa, el teatro de los siglos XVI-XVII, y a aclarar influencias y mensajes no fáciles de percibir sin su ayuda en las obras de artes plásticas.

Los libros de emblemas españoles eran prácticamente desconocidos cuando comenzamos a estudiarlos; por varias razones:

- La escasez de ejemplares en bibliotecas especializadas y su dispersión geográfica, lo que dificultaba un estudio del género teniendo en cuenta el conjunto.
- La inexistencia de ediciones modernas, salvo alguna facsimilar bastante descuidada.
- El hecho de que sea un género híbrido, en que aparece imagen, texto en latín –y en otras lenguas– y texto en español que requiere que quienes se ocupen del estudio atiendan a todas las partes de cada emblema. Un emblema es una composición mixta articulada por lo general en tres elementos (Fig. 1):
 - unas palabras que sintetizan el concepto,
 - una imagen que lo ilustra,
 - y unos versos que explican el sentido de la composición.

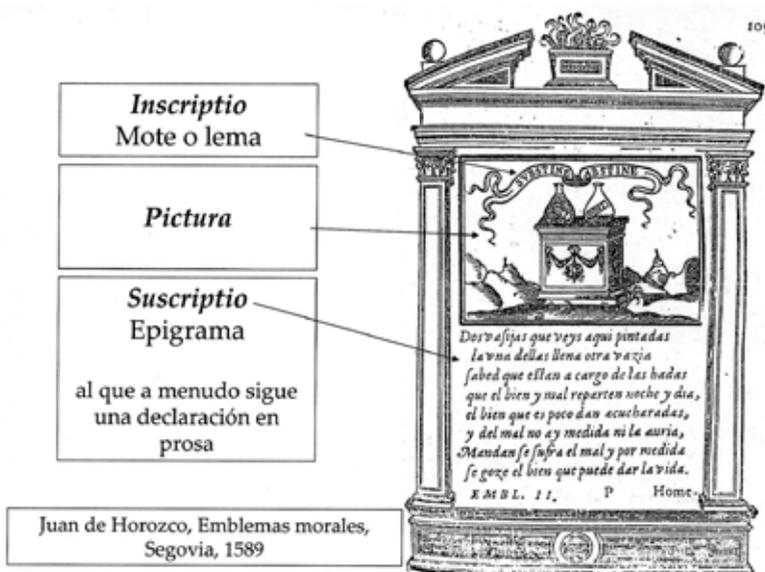


Fig. 1. Estructura retórica de un emblema.

El ejemplo que vemos en la figura 1 es un emblema del libro de Juan de Horozco *Emblemas morales* (1589), en el que podemos apreciar los componentes del *emblema triplex*, lo que nos da idea de la diversidad de materiales con los

que ha de trabajar el investigador, pues intervienen textos en latín, en español, en verso y en prosa, además de imágenes que ilustran el concepto en un grabado (xilográfico en este caso).

Para estudiar los libros de emblemas, además de ser precisa una buena formación en el tema, lo que comenzaba por un amplio conocimiento de la bibliografía internacional (conocer bien las ediciones españolas y otras fuentes primarias) y toda la bibliografía secundaria, se requería un trabajo conjunto de especialistas en Filología española, Filología clásica, Historia del Arte, Bibliografía analítica, Historia del libro y el grabado...

Para disponer de los libros de emblemas españoles (y de distintas ediciones si es que las hubiere) acometimos en 1993 la digitalización de microfílm, para poder acceder a libros de extremada rareza sin tener que realizar un enorme esfuerzo económico y físico, porque los originales estaban en bibliotecas dispersas y lejanas. Hoy esto nos puede parecer extraño, sobre todo a los jóvenes, por el milagro que ha supuesto Internet y la digitalización masiva de colecciones de libros. En los inicios de los años 90, la digitalización teníamos que hacerla con una enorme máquina Canon que costaba una fortuna.

Una vez que dispusimos de copias digitalizadas, los distintos equipos del grupo podían acometer el análisis de las distintas partes que por su especialización les habían sido encomendadas.

Fue preciso diseñar una serie de bases de datos y desarrollar aplicaciones *ad hoc* para alimentarlas con los elementos que considerábamos pertinentes en cada una de las partes de cada emblema. Eran bases de datos y aplicaciones locales, en portátiles o en ordenadores personales. No era tan fácil como hoy, que podemos tener aplicaciones de gestión en Internet (ver apartado 6.1 "Evolución en la arquitectura del sistema").

Sometimos a análisis casi 2.000 emblemas.

Los miembros del **equipo de Literatura española** realizaron una introducción de cada obra, detallando la edición que se había seleccionado, transcribieron los epigramas y analizaron el tipo de verso y de estrofa, redactaron un resumen de la glosa de cada emblema, realizaron *thesauros* de autoridades citadas, índices onomásticos y palabras clave, exempla utilizados... y algún otro aspecto que luego no incluimos en la BD final.

Los **especialistas en latín y griego** se ocuparon de los lemas o motes (que generalmente están en latín en estas obras), su transcripción, su traducción al español y la búsqueda de sus fuentes, que a menudo están en la literatura clásica.

Los especialistas en **Historia del Arte**, realizaron una descripción de los motivos pictóricos empleados en la *pictura* y el significado que el autor le da en el emblema analizado en concreto.

También se crearon unas bases de datos de **análisis retórico, estilístico y del tipo de técnicas de grabado empleadas**, pero finalmente descartamos incluir en la base de datos final esos estudios.

La implementación de las bases de datos, creación de las páginas web, diseños e implementación de las interfaces de consulta fueron trabajos realizados por un equipo informático dirigido por Nieves Rodríguez Brisaboa. Desde que acudí a ella en busca de ayuda como experta en bases de datos (1995), podemos decir que se produjo una interesante simbiosis, pues los proyectos iniciales permitieron la creación del Laboratorio de Bases de Datos de la UDC, la elaboración de varias tesis doctorales y el beneficio mutuo de los dos equipos. En 1996 fui-

mos seleccionadas ambas para presentar en el encuentro anual del CICYT nuestro proyecto coordinado⁶.

En esta etapa los miembros del SIELAE, que partíamos de una formación en Filología u otras ramas de Humanidades, tuvimos que complementarla con conocimientos que precisábamos sobre herramientas y métodos informáticos que no nos habían proporcionado nuestros estudios universitarios. Y sobre todo, nos enfrentábamos a la necesidad de tomar decisiones normativas respecto a cómo organizar el conocimiento, cómo ponernos en el lugar de quien deseaba buscar, realizar *thesauros* que ayudaran a sistematizar los contenidos, etc. Eso se hacía especialmente difícil en el estudio iconográfico de los emblemas, porque todavía no existía –y cuando existió resultó inaccesible para nosotros desde un punto de vista económico– el sistema de clasificación iconográfica Iconclass.

En esa primera fase, el SIELAE ofreció a través de Internet la posibilidad de consulta de 24 libros de emblemas españoles (de extrema rareza) en formato digital facsimilar, y añadía el fruto de análisis especializado de cada una de las partes de cada emblema: la transcripción y traducción de los motes, la transcripción de todos los epigramas, la posibilidad de buscar por estrofas o tipos de versos, los resúmenes de todas las glosas o declaraciones, la descripción de motivos representados en las *picturae* y su significado y una serie de *thesauros* para buscar autoridades citadas, palabras clave, onomástica, etc. La *Biblioteca Digital de Emblemática Hispánica* está integrada ahora en el portal BIDISO: <<http://www.bidiso.es/EmblematicaHispanica/>>.

Además, desde 1996, el SIELAE ofrecía a través de Internet la bibliografía más actualizada y completa sobre emblemática internacional en la página *Literatura Emblemática Hispánica. Bibliografía, enlaces y noticias* <<http://www.bidiso.es/emblematica>>, distinguiendo las entradas bibliográficas en los bloques *repositorios*, *ediciones* y *estudios*. En la actualidad contiene 2.850 fichas, que ahora pueden consultarse en el nuevo portal BIDISO, en el apartado de bibliografía especializada. Continúa siendo un recurso de referencia recomendado constantemente a sus miembros por *The Society for Emblem Studies*.

En 1999 se amplió el proyecto (*Biblioteca Digital de Libros de Emblemas Traducidos*), que pretendía analizar ocho libros de emblemas escritos en distintas lenguas que fueron traducidos al español en su tiempo. El proyecto analizó cinco, y no se completó. Lo realizado está accesible en el portal BIDISO:

<<http://www.bidiso.es/LiteraturaEmblematica/SelectLanguage.do?language=es>>

Nuestro proyecto inicial hoy sigue dando un importante servicio, a pesar de que, lógicamente, tiene sus limitaciones, fundamentalmente ligadas a la calidad de las imágenes digitales, no procedentes, como los posteriores, de la digitalización directa de los libros.

En las dos primeras bibliotecas:

- Base de datos y biblioteca digital de los libros de emblemas españoles de los siglos XVI y XVII,
- Biblioteca Digital de Libros de Emblemas Traducidos,

⁶ López Poza, Sagrario (*et al.*), "Acceso a una base de datos de literatura emblemática a través de Internet", en *Red Iris. Boletín de la Red Nacional de I + D RedIRIS*, 38 (diciembre 1996), pp. 55-63. <<http://www.rediris.es/difusion/publicaciones/boletin/38/ponencia6.html>>.

se ofrece la posibilidad de hacer búsquedas complejas como por ejemplo en cuáles de los libros aparece una imagen que representa a un caracol, una cigüeña, o cualquier otro motivo iconográfico. Pueden hacerse búsquedas complejas sobre los motes, sobre cualquier palabra que aparezca en los epigramas, en el resumen de la glosa, por palabras clave y significados de los emblemas...

Los datos siguen siendo de interés y de mucha ayuda, y aunque la calidad de la imagen facsimilar es, por decirlo así, mejorable, cumple perfectamente su finalidad de ayuda al estudio, y en caso de que se precise una mejor resolución, hoy puede suplirse enseguida acudiendo a las muchas ediciones buenas facsimilares de los libros, realizadas sobre todo en la segunda parte de la primera década del siglo XXI.

Los mejores proyectos que han ido surgiendo, a partir de 2003 sobre este tema, han aprovechado nuestra experiencia y han aportado todas las ventajas que el tratamiento de imagen y los lenguajes de marcado han ido aportando en los últimos años.

Considero ejemplar en ese sentido el *Emblem Project Utrecht* (y el trabajo realizado por Else Stronk y Peter Boot), sobre libros de emblemas holandeses, pionero en la aplicación de la TEI en las transcripciones de los textos y en el uso de Iconclass, y los proyectos de la Glasgow University Emblem Website (Alison Adams para los libros de emblemas franceses y Stephen Rawles para las principales ediciones de los emblemas de Andrea Alciato, además de otros importantes colaboradores). Estos proyectos son un modelo de lo que la Informática y las Humanidades pueden hacer unidas. También es digno de mención el proyecto que dirige Mara Wade en la universidad de Illinois sobre libros de emblemas alemanes, con la participación de la biblioteca Herzog August Bibliothek. Ese proyecto de portal tenía la intención de reunir (recolectar) para hacer accesible a través de él otros proyectos de libros de emblemas europeos.

- Emblem project Utrecht
<http://emblems.let.uu.nl/index.html>
- French Emblems at Glasgow
<http://www.emblems.arts.gla.ac.uk/french/>
- Alciato at Glasgow
<http://www.emblems.arts.gla.ac.uk/alciato/>
- Open Emblem Portal (University of Illinois y Herzog August Bibliothek)
<http://emblematica.grainger.illinois.edu/OEBP/UI/SearchForm>

Con el inicio del siglo XXI, las bibliotecas digitalizaron masivamente libros antiguos con fines de preservación y a la vez para difundir el patrimonio en Internet (a veces en colaboración con Google) y se produjo un gran impulso cuando la Unión Europea estimuló –y dio importantes subvenciones para ello– la digitalización del patrimonio bibliográfico.

El SIELAE consideró que debía aprovechar sus conocimientos de expertos en emblemática para ofrecer a otros investigadores y estudiantes la posibilidad de encontrar rápidamente en la web las ediciones de libros de emblemas, sin tener que pasar horas buscando ediciones diversas, etc. Así, desde 2006, ofrecemos un *Catálogo de ediciones digitales de libros de emblemas y obras afines acce-*

sibles en Internet/**DEBOW** (Digital Emblem Books on Web): <<http://www.bidiso.es/emblematica/CatalogoDEBOW.pdf>> que va por la quinta edición (julio 2013) y que contiene 1.266 fichas de ediciones de libros de emblemas u obras afines en formato digital, con uno o más ejemplares disponibles en Internet, y los enlaces correspondientes para un acceso directo. El número de ejemplares de los que proporcionamos enlace es de 1.934. El fichero en PDF ocupa 191 páginas.

Esta es una característica del SIELAE que quiero destacar. Desde el principio, la prehistoria de Internet entre los grupos españoles, nos dimos cuenta de que TODO lo que produjéramos con dinero público había de servir para impulsar el conocimiento de otros investigadores o público interesado. Esto, que hoy puede parecer una obviedad, era extremadamente raro en los años 90. De hecho, los colegas me preguntaban por qué regalábamos nuestro conocimiento. Muchos tardaron bastantes años en darse cuenta de que para la investigación, tan importante o más que los derechos intelectuales es la difusión. Hay colegas que han pasado diez años trabajando en bases de datos que han publicado en CD-ROM con una editorial comercial. Esa editorial tiene "secuestrado" ese trabajo por diez años, y no pueden sus creadores difundirlo en Internet, con lo cual es casi opaco y como el producto comercial es muy caro, poca gente lo conoce y lo cita.

El portal BIDISO <<http://www.bidiso.es>> integra los recursos y colecciones creados durante 20 años por los miembros del SIELAE, que ha logrado ir manteniendo las colecciones antiguas e integrar otras nuevas. Lo forman cuatro bloques (uno integra las dos colecciones de libros de emblemas), otro las Relaciones de sucesos (del que hablará Nieves Pena), otro los inventarios de bibliotecas del siglo de Oro (del que hablará Carlota Fernández Travieso) y otro de Polianteas.

2.2. Las polianteas

Las polianteas son enciclopedias o repertorios de citas y erudición diversa que surgieron a comienzos del siglo XVI⁷ y que fueron muy utilizadas por poetas, oradores, predicadores, pintores y artistas en general como ayuda para elaborar sus producciones. Bajo títulos como *Polyanthea*, *Officina*, *Sylva*, *Hortus floridus*, *Thesaurus*, *Theatrum*, *Syntaxis*, *Panoptikon*, *Argumenta*, o sus equivalentes en lenguas vulgares: *Teatro*, *Fábrica*, *Jardín*, *Florilegio*, etc. las polianteas reúnen sentencias de los clásicos, adagios, apotegmas, proverbios, anécdotas de la Historia Sagrada o profana, exempla, fábulas; proporcionan definiciones, clasificaciones de conceptos, citas, exempla, símiles, etc., un material que servía para adornar de erudición el discurso o como fuente de invención. En general permitían hacer ostentación de una erudición que la mayor parte de las ocasiones no era auténtica, sino tomada de este tipo de herramientas auxiliares, y por ello se convirtieron en instrumentos imprescindibles para todo hombre

⁷ La *Polyanthea* del humanista italiano Domenico Nani Mirabelli (*Polyanthea: opus suavis-simis floribus exornatum compositum per Dominicum Nanum Mirabellium [...] impressum per magistrum Franciscum de Silua in incluta urbe Saonae: impensa integerrimi viri ac [...] Bernardini de Ecclesia [...], idibus februaryiis 1503 [13 febrero 1503]*) obtuvo un éxito inmediato y le siguieron, al menos, 26 ediciones hasta 1681 en versiones que fueron ampliándose con diversas participaciones, entre las que destacó la de un profesor de origen alsaciano que enseñó matemáticas y griego en Friburgo, Joseph Lange, con el que colaboró un tal François Dubois de Lille a partir de la edición de 1621. Para más detalles, véase López Poza (2000).

de letras (sagradas o profanas) y se produjeron hasta finales del siglo XVII. En cierto sentido, podríamos decir que eran algo así como el *Google* y la *Wikipedia* del Siglo de Oro.

En realidad eran producciones impulsadas por el florecimiento de la imprenta (pues los impresores se percataron pronto de que podía ser un producto de éxito en las ventas) al facilitar lo que había sido y seguía siendo una práctica individual que era un procedimiento de extracción de citas de las lecturas personales, que cada individuo recogía en un cuaderno dividido en 12 ó 14 secciones de lo que denominaban *loci communes*. Los humanistas (tenemos ejemplos detallados de Luis Vives, Erasmo, Miguel de Salinas, Lorenzo Palmireno, Justo Lipsio y otros) explican con detalle cómo preparar el propio cartapacio personal de citas o *codex excerptorius* (donde el estudiante debía anotar cuanto de sus lecturas le parecía que podía serle de utilidad ulterior). La necesidad por una parte, y el interés comercial por otra, propiciaron que, ya desde comienzos del siglo XVI, con la publicación de la *Polyanthea* de Domenico Nani Mirabelli (Savona, Italia, 1503) organizada por orden alfabético, se publicaran bastantes obras de este tipo. Los procedimientos de organización del conocimiento son diversos, intentando reflejar el mundo en taxonomías y criterios diferentes. Los argumentos eruditos extraídos de toda clase de saber se hicieron imprescindibles en el discurso de cualquier escrito culto de los siglos XVI y XVII y los predicadores, oradores y poetas tenían siempre en su biblioteca una de estas obras, casi todas escritas en latín.

Tan frecuente era el uso de estas obras, que algunos autores se quejan de que ya no se acude a la lectura directa de los clásicos, sino a estas recopilaciones.

Para quienes estudiamos la producción literaria, iconográfica, cultural, de los siglos XVI al XVIII, este tipo de obras resulta de enorme utilidad: para ediciones anotadas, o para comprender programas iconográficos, o estudios diversos sobre Literatura, Arte o transmisión y organización del conocimiento. Este patrimonio cultural europeo suele conservarse en bibliotecas con fondo antiguo, pero no resulta fácilmente accesible. Por esa razón comenzamos a digitalizar este tipo de obras y a hacer índices en español de conceptos recopilados en algunas de las principales ediciones, con enlaces a las páginas en concreto donde se tratara del asunto buscado.

Con el tiempo, vimos que en diversas bibliotecas con fondo antiguo, al digitalizar sus materiales, lo habían hecho también de las polianteadas, y entonces incorporamos fichas en nuestro catálogo de esas ediciones o ejemplares y pusimos enlaces que llevaban directamente a ellos.

A través de nuestro portal BIDISO, ofrecemos acceso a 50 ediciones de obras digitalizadas de este género (bien realizada por nosotros o incluyendo fichas con un enlace a los ejemplares digitalizados en otras bibliotecas); se ofrece también bibliografía secundaria (repertorios y estudios) y enlaces a páginas de interés sobre el tema.

3. PRESENTACIÓN DEL LABORATORIO DE BASES DE DATOS

El LBD: Laboratorio de Bases de Datos (<http://lbd.udc.es>) es un grupo de investigación ubicado en la Facultad de Informática de la Universidade da Coruña formado por profesorado del área de Lenguajes y Sistemas Informáticos. Este

grupo, coordinado por la catedrática Nieves Rodríguez Brisaboa, tiene en las *Bibliotecas Digitales* y la *Recuperación de información* dos de sus líneas de investigación principales.

Así, desde su creación, ha trabajado en equipos multidisciplinares, siendo responsable de proyectos como la *Biblioteca Virtual Gallega*⁸ (Places *et al.*, 2007a), la *Hemeroteca Virtual de la Real Academia Gallega*⁹, o el *Centro de Documentación de la Asociación de Escritores en Lengua Gallega* (AELG)¹⁰.

Como ya se ha comentado, uno de los trabajos en los que el LBD ha estado más implicado desde su creación, ha sido el que corresponde al SIELAE: *Seminario Interdisciplinar para el Estudio de la Literatura Áurea Española* (<http://www.bidiso.es/SIELAE>), dirigido por la profesora Sagrario López Poza, y del que ha surgido la *Biblioteca Digital Siglo de Oro* (BIDISO), que es el centro de la comunicación que nos ocupa. Ambos equipos colaboran con esfuerzos individuales y conjuntos en las Bibliotecas que forman BIDISO.

4. METODOLOGÍA DE TRABAJO

La metodología que utilizamos en el desarrollo de las Bibliotecas Digitales que forman BIDISO, está basada en la ya clásica metodología en cascada modificada con iteraciones, donde en cada iteración se desarrollan las tareas de análisis, diseño, implementación y pruebas. Guiados por esta metodología, las fases que realizamos en cualquier proyecto son, fundamentalmente, las siguientes:

- **Definición de objetivos:** se define claramente que se quiere conseguir con la biblioteca o proyecto que se va a comenzar. Esta tarea es realizada íntegramente por el SIELAE, que cuenta con el asesoramiento esporádico del equipo de informática.
- **Análisis:** se definen los requisitos funcionales y no funcionales que debe cumplir la aplicación, se explicitan las necesidades de almacenamiento de información y se formalizan en un modelo conceptual. Esta tarea es llevada a cabo de forma conjunta por los dos equipos. De hecho, el éxito de esta fase depende de lo fructífera que sea la colaboración.
- **Diseño:** se divide en diseño arquitectónico, diseño de bajo nivel y diseño de interfaces de usuario. Esta fase es llevada a cabo fundamentalmente por el LBD, con las pertinentes indicaciones del SIELAE en lo que se refiere a las interfaces de usuario.
- **Implementación:** se trata de la construcción de las aplicaciones en sí. El LBD es el responsable de esta tarea.
- **Pruebas:** se realizarán pruebas unitarias, de integración y, las más importantes, de aceptación de usuario final. El SIELAE está involucrado en las pruebas de aceptación final.
- **Implantación:** se instala la nueva biblioteca digital o, en general, la nueva aplicación desarrollada, en un entorno de producción y accesible desde una URL pública. Esta tarea es realizada por el LBD.

⁸ <<http://bvg.udc.es>>.

⁹ <<http://lbd.udc.es/RAG-2004-2012>>.

¹⁰ <<http://www.aelg.org/Centrodoc/>>.

Aparte de las tareas que dicta la metodología, existen otras dos que comienzan después de la implantación y se alargan durante toda la vida de la Biblioteca Digital. Se trata de la alimentación de la BD, de la que se encarga el equipo de Filología, y el mantenimiento correctivo y evolutivo, que ejecuta el LBD, en la mayor parte de las ocasiones, a instancia del SIELAE.

5. ASPECTOS CLAVE EN EL DESARROLLO DE BIDISO

En el desarrollo de BIDISO se han tenido en cuenta algunos aspectos críticos que consideramos fundamentales para el éxito de cualquier Biblioteca Digital. Nos referimos a:

- Diseño de datos: almacenamiento y búsqueda.
- Autonomía y facilidad de gestión.
- Interfaces de usuario: adaptadas y accesibles.
- Integración e intercambio de datos.

A continuación se describen estos aspectos en detalle.

5.1. Diseño de datos: almacenamiento y búsqueda

El objetivo del diseño de datos en una Biblioteca Digital es almacenar la información que sea relevante para el dominio de que se trate, pero también es importante llevarlo a cabo de manera que posteriormente la información almacenada se pueda explotar de forma eficiente a través de consultas e informes.

Por este motivo, la fase de modelado o diseño de datos es una fase crucial en el desarrollo de una Biblioteca Digital ya que de ella dependerá el soporte que la Biblioteca dé a estudios posteriores.

5.2. Autonomía y facilidad de gestión

Los responsables de los contenidos deben tener la posibilidad de gestionar de forma autónoma toda la información de la base de datos. Además, esta gestión debe poder hacerse de forma colaborativa por el conjunto de personas del equipo de trabajo, teniendo en cuenta que cada persona podrá tener una función y unas responsabilidades específicas a realizar y/o unas restricciones y privilegios determinados para acceder a la información ya almacenada.

Así, cada una de las bibliotecas digitales integradas en BIDISO cuenta en la actualidad, con su propia herramienta web de gestión de acceso restringido, que permite que diferentes perfiles (colaboradores y administradores, en general) con diferentes tareas y responsabilidades puedan encargarse de la gestión de los datos de forma autónoma y sencilla (Cotelo *et al.*, 2005; Places *et al.*, 2007b).

5.3. Interfaces de usuario: adaptadas y accesibles

Aunque, como dijimos, el diseño de la base de datos es un aspecto crítico en el éxito de una Biblioteca Digital, también lo es la interfaz que la biblioteca proporcione para consultar los datos que tenga almacenados, ya que es la herramienta que usuarios, expertos o no en el dominio y en tecnologías informáticas, utilizarán para acceder a la información.

Así, la estructura web de la Biblioteca debe diseñarse para facilitar la navegación a través de los datos almacenados, y sus interfaces de usuario deben permitir realizar consultas sencillas a la base de datos, pero también consultas más complejas, construidas a través de combinaciones de criterios, que exigirán el desarrollo de interfaces de usuario también más complicadas (Brisaboa *et al.*, 2001).

5.4. Integración e intercambio de datos

Los sistemas de integración e intercambio de datos más extendidos en la actualidad, son los llamados débilmente acoplados. Este tipo de sistemas permite integrar bibliotecas heterogéneas, al hacerlo a través de atributos claramente comunes, en vez de tratar de encajar los modelos de datos de las bibliotecas que normalmente son muy complejos. Además, los sistemas de integración débilmente acoplados suelen encargarse de la fase de consulta, pero delegan en cada Biblioteca Digital de la que se obtenga algún resultado de búsqueda, la presentación de las respuestas.

Los estándares que están ayudando a favorecer la implantación de los sistemas integrados débilmente acoplados y que fomentan el intercambio de información son: el *Protocol for Metadata Harvesting* de la *Open Archives Initiative* (OAI-PMH)¹¹ y *Dublin Core Metadata Initiative* (DC)¹². También la *Text Encoding Initiative* (TEI)¹³ está proporcionando avances en este sentido. Su base XML permite que tenga un formato ideal para el intercambio de información, pero además resulta más rico que DC, ya que permite marcar aspectos del contenido de los documentos, además de sus metadatos.

Una de las bibliotecas de BIDISO, concretamente, la Biblioteca Digital de Libros de Emblemas Traducidos cuenta con la funcionalidad de exportar sus datos a un esquema TEI definido para bases de datos y bibliotecas digitales de Literatura Emblemática. Por otro lado, SIELAE forma parte de "Aracne: Red de Humanidades Digitales y Letras Hispánicas"¹⁴, que tiene como parte de sus objetivos, el construir un sistema de consulta integrado por las bibliotecas participantes en la red, basado en OAI-PMH, del que BIDISO formará parte como *Data Provider*. El LBD será el encargado del diseño y desarrollo del sistema completo.

5.5. Mantenimiento evolutivo

El mantenimiento evolutivo está formado por aquellos desarrollos que es necesario realizar para acomodar la Biblioteca a las nuevas tecnologías que se van implantando en el mercado, como nuevas versiones de navegadores web, y a los nuevos dispositivos que van apareciendo, por ejemplo, *tablets* e incluso, teléfonos móviles *smartphone*. Sin embargo, también puede considerarse mantenimiento evolutivo la implementación de nuevas funcionalidades que adapten la Biblioteca a las nuevas expectativas de los usuarios.

Afrontar el mantenimiento evolutivo de una Biblioteca Digital es imprescindible para que la biblioteca se mantenga útil en el tiempo. Se trata de un trabajo que

¹¹ <<http://www.openarchives.org/pmh/>>.

¹² <<http://dublincore.org/>>.

¹³ <<http://www.tei-c.org/index.xml>>.

¹⁴ Véase en este mismo volumen el trabajo presentado por Consolación Baranda y Eduardo Rodríguez.

tiene un coste elevado pero que muchas veces no supone ninguna mejora real en la biblioteca, sino que es necesario hacerlo para que todo siga funcionando como lo venía haciendo. Además, no muchas subvenciones incluyen entre los gastos subvencionables el mantenimiento del software, por lo que, a pesar de ser un factor tan crítico, se le presta poca atención.

6. EVOLUCIÓN DE BIDISO

BIDISO nació en 1996 cuando se creó la primera de sus Bibliotecas Digitales, la Biblioteca Digital de Literatura Emblemática Hispánica (Paramá *et al.*, 2006). Aquellos eran los inicios de BIDISO, pero también de la propia web, que no había hecho más que empezar a andar y que evolucionaría y mejoraría con los años.

En los siguientes apartados, se explica la evolución de BIDISO, desde su arquitectura, hasta los aspectos de interfaz de usuario relacionados con los navegadores, pasando por los lenguajes y tecnologías de programación.

6.1. Evolución en la arquitectura del sistema

El cambio más importante que han registrado las Bibliotecas Digitales que forman BIDISO ha sido el que se refiere a la arquitectura del sistema. Como ya se adelantó en apartados anteriores, en los años en los que las primeras Bibliotecas (Paramá *et al.*, 2006) fueron creadas (1996) y posteriores, Internet no estaba en absoluto extendida, y mucho menos había posibilidad alguna de conexión en Bibliotecas, Archivos, etc. en los que los investigadores del equipo de Filología digitalizaban los documentos y recopilaban información de catalogación. Por lo tanto, fue necesario desarrollar aplicaciones locales de gestión de datos e instalarlas en los portátiles de los investigadores para que pudieran usarlas de manera *standalone*.

La información de las bases de datos locales era volcada de forma periódica a la base de datos central a través de un proceso muchas veces más manual que automático, ya que aunque existían protocolos de codificación para las diferentes entidades de información (obras, ediciones, ejemplares, etc.), un pequeño error suponía, por ejemplo, no detectar de forma automática que dos ejemplares catalogados por dos investigadores diferentes pertenecían a la misma edición, con lo cual era imprescindible hacer una revisión manual.

En la figura 2 se presenta la arquitectura inicial que acabamos de describir, y que derivó, con la expansión de Internet, en la arquitectura actual de la figura 3. En esta nueva arquitectura, la gestión de la información se realiza, de forma colaborativa por todos los miembros del equipo de trabajo, a través de una única aplicación de gestión que se ejecuta sobre una única base de datos y un repositorio también centralizado. Con esta arquitectura se consigue, por un lado, evitar la redundancia de información que permitía la arquitectura anterior, y, por otro lado, la información añadida o las modificaciones sobre los datos existentes quedan inmediatamente disponibles para consulta.

6.2. Evolución en los lenguajes de programación y los sistemas gestores de bases de datos

El lenguaje de programación que usamos para las aplicaciones locales de los portátiles de los investigadores era fundamentalmente Delphi. Las aplica-

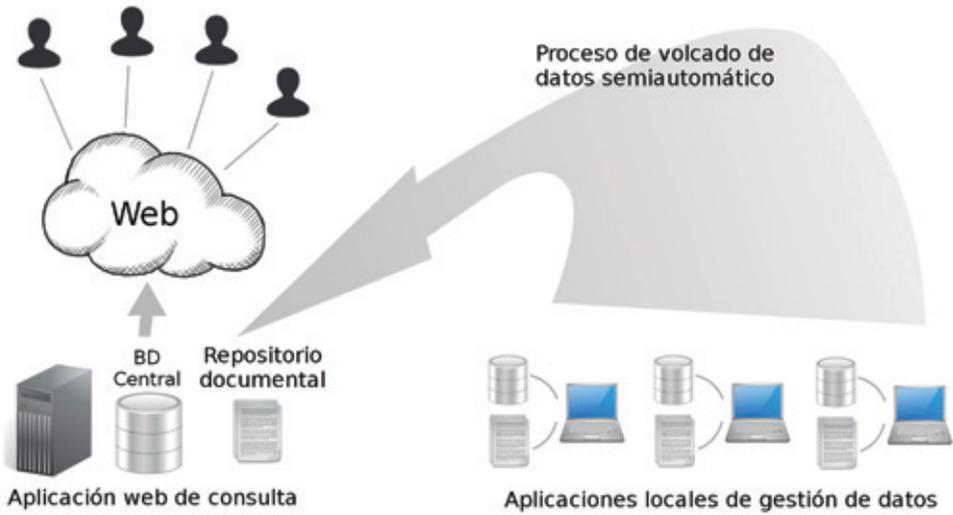


Fig. 2. Arquitectura de sistemas inicial.



Fig. 3. Arquitectura de sistemas actual.

ciones así programadas se ejecutaban contra una base de datos implementada en Paradox. También hicimos algún desarrollo en FoxPro, aunque fueron muy esporádicos. Como ya comentamos en el apartado anterior, estas aplicaciones desaparecieron, por evolución a una arquitectura con una única base de datos y repositorio central.

Los lenguajes de programación con los que construimos las aplicaciones de servidor, sí fueron variando a lo largo de los años a medida que la tecnología de desarrollo de aplicaciones web maduraba. Las primeras versiones de las Bibliotecas Digitales fueron programadas con tecnología CGI (*Common Gateway Interface*), siendo el lenguaje de programación usado C. La programación se realizaba en el entorno visual Saphire Web.

Los primeros usos de JAVA permitieron construir aplicaciones que se ejecutaban del lado del cliente, a través de Applets. Esta tecnología, que permitía construir aplicaciones web muy portables, y con la que construimos la segunda versión de la Biblioteca Digital de Literatura Emblemática Hispánica para sustituir a la implementada con HTML y CGI, tuvo que ser desechada porque requería que los usuarios tuviesen bien configurados sus navegadores para que todo funcionase correctamente, lo cual era muchas veces difícil porque el nivel de conocimientos de informática de los usuarios de la Biblioteca no les permitía realizar esta tarea con facilidad.

Finalmente, las aplicaciones Java derivaron en la tecnología JSP, que al igual que CGI permite construir aplicaciones del lado del servidor. Las versiones actuales de nuestras Bibliotecas están construidas con esta tecnología (ver apartado 7.1), y usan JavaScript para la lógica en el cliente.

En cuanto a los Sistemas Gestores de Bases de Datos (SGBDs), los primeros implantados en el servidor, fueron dos sistemas propietarios: primero Informix, y más tarde SQLServer. En la época en que las Bibliotecas de BIDISO estaban soportadas por estos SGBDs, el pagar las licencias de estas grandes marcas era la única forma de garantizar un buen rendimiento en la base de datos. Más tarde, con la aparición y sucesivas mejoras de MySQL y PostgreSQL, dos SGBD libres, migramos las bases de datos que quedaron finalmente implantadas en estos gestores.

6.3. Evolución en los navegadores

A lo largo de todos estos años, hemos asistido a una evolución continua de los navegadores (a medida que también evolucionaba la propia web, con nuevos protocolos, lenguajes y metodologías): han surgido unos nuevos y otros han desaparecido, y todos han ido publicando nuevas versiones que ampliaban sus funcionalidades, proyectaban otro aspecto, pero en todo caso, cambiaban sus interfaces con las aplicaciones que se ejecutaban en su entorno, y no siempre de la misma forma, haciendo que estas últimas tuviesen que evolucionar a su vez teniendo en cuenta las características especiales de cada uno de ellos.

Así, BIDISO nació con el Internet Explorer, en su versión 3, en el que se ejecutaba la aplicación de consulta de la Biblioteca Digital de Literatura Emblemática Hispánica y ha tenido que adaptarse a todas las siguientes versiones hasta la 10 actual; ha visto triunfar y desaparecer el Netscape Navigator; y se ha adaptado a los no ya tan recientes, Mozilla Firefox, Google Chrome y Apple Safari, que son los más usados en la actualidad.

Hoy en día, quizás se pueda afirmar que la web es más "estable", pero sobre todo, el grado de estandarización es más alto, por lo que las adaptaciones de

las aplicaciones a las diferentes versiones de los navegadores se soportan de forma más sencilla. Sin embargo, esto sirve únicamente para el ámbito de los ordenadores personales, ya que en el ámbito de los dispositivos móviles (*tablets* y *smart phones*) estamos en plena explosión con numerosas propuestas de hardware, resoluciones de pantallas diferentes, navegadores diferentes, etc. y es justo a donde pronto tendrán que evolucionar las Bibliotecas Digitales, por ser estos dispositivos móviles el hardware a través del cual los usuarios, cada vez más, consumen contenidos digitales.

7. ARQUITECTURA, TECNOLOGÍA E INFRAESTRUCTURAS ACTUALES

7.1. Entorno de desarrollo

Las aplicaciones web que forman BIDISO están basadas en una arquitectura en capas que proporciona independencia entre el acceso a datos, la lógica de la aplicación/negocio y la vista (un esquema de esta arquitectura se puede ver en la Fig. 4). Esto, unido al uso sistemático de patrones de diseño en su cons-

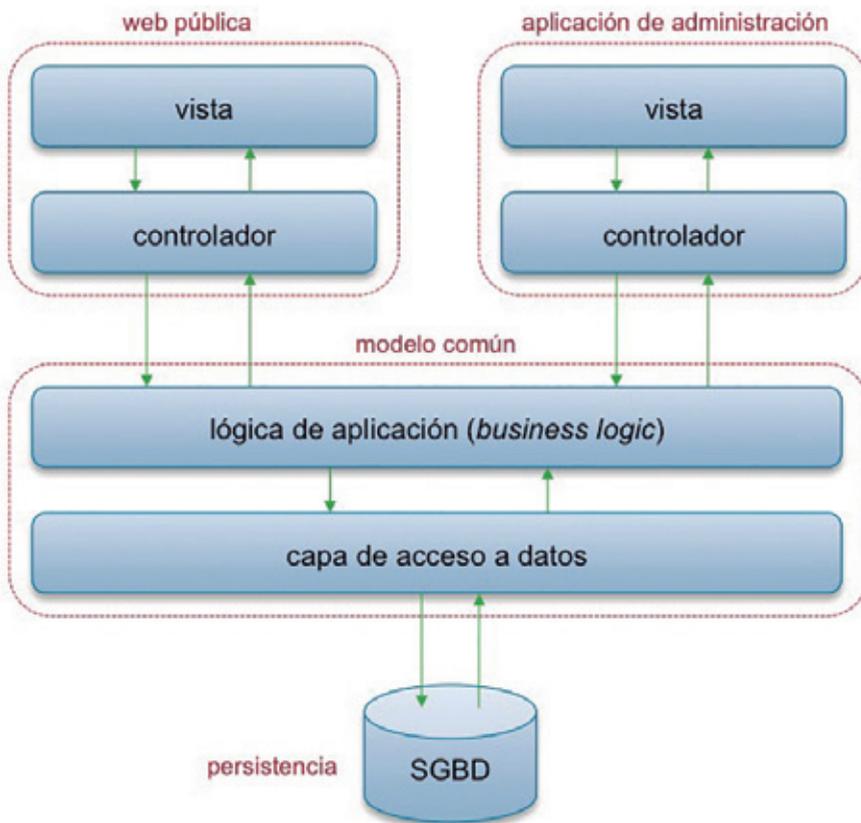


Fig. 4. Arquitectura de aplicaciones.

trucción, supone una garantía para asegurar la robustez y escalabilidad de las aplicaciones construidas.

Para la programación, usamos *Java 2 Enterprise Edition (J2EE)* que está diseñado utilizando estándares abiertos que permiten la utilización de diferentes tecnologías en la implementación de los componentes de las distintas capas. En concreto, las tecnologías que utilizamos en la implementación de cada capa son las siguientes:

- **Capa de datos:** el Sistema Gestor de Bases de Datos (SGBD) elegido es *PostgreSQL*, y el entorno de desarrollo *Hibernate*.
- **Lógica del sistema:** utilizamos *Spring Framework*, en particular, las librerías de *Spring MVC*, que implementan el patrón arquitectónico *Model-View-Controller*, *Spring Security* que facilita a autenticación y control de permisos en las aplicaciones, o el módulo de inversión de control que permite reducir el acoplamiento entre los distintos componentes, para mejorar su integración.
- **Capa vista:** utilizamos *JavaServer Pages (JSP)* para la creación de las páginas web que permiten presentar la información.

Además, para apoyar el proceso de desarrollo, utilizamos herramientas de uso libre o de construcción propia que nos facilitan la organización de la documentación relacionada con el proyecto, la comunicación entre los diferentes miembros del equipo de desarrollo y el control en el seguimiento de la planificación. Son de destacar: *Eclipse* y *CVS* para el código fuente y *SVN* para la documentación del proyecto. Tenemos un único repositorio *CVS* y 4 repositorios *SVN* que alcanzan ya los 284 MBytes de información; *Xestiona*, para el control de la planificación y el seguimiento del proyecto; *Redmine*, para el control de incidencias, etcétera.

7.2. Entorno de producción

En la actualidad, BIDISO está formado por más de 25 aplicaciones web y páginas estáticas (páginas web construidas directamente en HTML), que están accesibles a través de dos dominios: <bidiso.es> y <rosalia.dc.fi.udc.es>.

Las aplicaciones se ejecutan en varias plataformas software: Apache, para las páginas estáticas, y Tomcat 5.5 y Tomcat 6.0 para las aplicaciones web. Los gestores de bases de datos (SGBD) que se usan en la actualidad son MySQL 5.5, para las más antiguas, y PostgreSQL 8.4, las más recientes. En los dos casos, se trata de SGBDs libres, que han demostrado tener unas buenas propiedades de rendimiento y seguridad.

El conjunto de aplicaciones que forma BIDISO está instalado en un servidor de producción HP DL180 G6 situado en el CECAFI, el Centro de Cálculo de la Facultad de Informática (Universidade da A Coruña). Este servidor está dotado de un procesador Intel Xeon E5606, 16GB de RAM y discos SATA de 500GB *hot plug*. El sistema operativo que tiene instalado es un Windows 2008 Server R2 Standard de 64 bits. El servidor está conectado a la red de la Universidad, con un ancho de banda de 100Mbps.

Soportando este servidor tenemos un SAI (Sistema de Alimentación Ininterrumpida) con capacidad para mantenerlo encendido durante 45 minutos desde que se activa, es decir, desde que se interrumpe el suministro eléctrico. A su vez, este SAI está detrás de un grupo electrógeno con una capacidad de 8 horas,

consiguiendo, por lo tanto, una autonomía total de 8 horas y 45 min en caso de fallo eléctrico.

7.3. Monitorización y plan de copias de seguridad

Como ya se ha comentado, el *hosting* de las aplicaciones que componen BIDISO se está realizando en servidores propios conectados a la red de la Universidade da Coruña. El mantenimiento del servidor y el software de base (Sistema operativo, antivirus, etc.), la monitorización del funcionamiento de las aplicaciones y las copias de seguridad de código, documentación, bases de datos y aplicaciones es realizada por los responsables de sistemas del Laboratorio de Bases de Datos.

Así, este equipo ha desarrollado y tiene implantada una herramienta que nos permite controlar de forma automática que las aplicaciones que forman BIDISO estén en ejecución. Esta herramienta de monitorización consulta periódicamente (actualmente, está programada para hacerlo cada 10 min) la página principal de cada aplicación. Si la respuesta es diferente de un código HTTP 200 (que indicaría que la aplicación está funcionando), envía un correo de alerta a los responsables de sistemas.

Por otro lado, hemos diseñado e implantado un plan de copias de seguridad que incluye un *backup* completo de todos los archivos (aplicaciones y repositorios CVS y SVN) el día 1 de cada mes, y un *backup* diferencial de archivos diario (en el *backup* diferencial, se almacenan únicamente los cambios que se produjeron desde el *backup* diferencial anterior). El plan de copias de seguridad de las bases de datos es independiente, e incluye *backups* diarios completos.

8. CONCLUSIONES

Durante dos décadas, el SIELAE y el LBD han trabajado juntos, han explorado las posibilidades de relación entre las Humanidades y la Informática y han demostrado que esa colaboración interdisciplinar puede ser muy fructífera. El trabajo realizado ha enriquecido mutuamente a los miembros de los dos grupos, ampliando sus conocimientos y enfoques y proyectando hacia el futuro actuaciones comunes y complementarias. Esa colaboración ha producido importantes realizaciones (publicaciones científicas, tesis doctorales o tesinas, materiales para el estudio y la investigación de estudiantes e investigadores, así como consecución de proyectos subvencionados), no sólo al equipo de Filología, sino también al equipo de informática. Así, el LBD ha conseguido financiación para desarrollar cuatro proyectos relacionados con BIDISO, dos de ámbito estatal y otros dos de ámbito autonómico. Los resultados de investigación obtenidos merecieron ser publicados en 6 revistas, 1 de ellas indexada en el *Journal Citation Reports* (JCR), y 7 congresos internacionales de reconocido prestigio en su área (Lenguajes y Sistemas Informáticos), además de en otros congresos iberoamericanos y españoles. El trabajo realizado también sirvió para mejorar la formación y especializar a un número representativo de alumnos y alumnas de informática: se han realizado 2 tesis de doctorado, una de ellas la correspondiente a la segunda autora de este texto (Ángeles Saavedra Places), 3 tesinas de licenciatura, y un total de 10 proyectos de fin de carrera de diferentes titulaciones de informática.

Los resultados del trabajo de estos dos grupos se ofrecen a la comunidad científica de manera abierta y gratuita, de modo que pueden servir para impulsar ulteriores investigaciones de especialistas o estudiantes de Filología Española (Literatura), Historia del Arte, Historia del periodismo, Filología Clásica, Bibliografía e Informática, lo que asegura que el dinero público recibido por los grupos, revierte en un beneficio para la sociedad.

BIBLIOGRAFÍA

- Brisaboa, N.R.; Penabad, M.R.; Places, A.S. y Rodríguez, F.J., "Tools for the Design of User Friendly Web Applications", en *Proceedings of the 2nd International Conference on Electronic Commerce and Web Technologies (EC-WEB 2001) - LNCS 2115*, Munich (Alemania), Springer-Verlag, 2001, pp. 29-38.
- Cotelo, J.A.; Paramá, J.R.; Penabad, M.R.; Places, A.S. y Vázquez, E., "Un sistema de administración de flujo de trabajo de creación de repositorios documentales para bibliotecas digitales", en *Actas de las V Jornadas de Bibliotecas Digitales (JBIDI 2005)*, Granada (España), Thomson, 2005, pp. 71-78.
- López Poza, Sagrario, "Polianteas y otros repertorios de utilidad para la edición de textos del Siglo de Oro", *La Perinola*, 4 (2000), pp. 191-207.
- Paramá, J.R.; Places, A.S.; Brisaboa, N.R. y Penabad, M.R., "The Desing of a Virtual Library of Emblem Books", en *Software - Practice and Experience*, 36(5), New York (Estados Unidos), John Willey & Sons, 2006, pp. 473-494.
- Places, A.S.; Brisaboa, N.R.; Fariña, A.; Luaces, M.R.; Paramá, J.R. y Penabad, M.R., "The Galician Virtual Library", en *Online Information Review*, 31(3), Bradford (Reino Unido), Emerald Group Publishing Limited, 2007, pp. 333-352.
- Places, A.S.; Brisaboa, N.R.; Paramá, J.R.; Pedreira, O. y Seco, D., "Managing the workflow of massive feeding of digital libraries", en *Research in Computer Science*, 32, Mexico (México), 2007, pp. 352-362.



RESUMEN

El SIELAE (Seminario Interdisciplinar para el Estudio de la Literatura Áurea Española) <<http://www.bidiso.es/sielae>> de la Universidade da Coruña (España) ha producido desde 1996 recursos digitales (bases de datos y bibliotecas digitales) en torno a cuatro áreas de investigación relacionadas con los siglos XVI y XVII: Emblemática, Relaciones de Sucesos, Polianteas e Inventarios de bibliotecas. Esos recursos fueron ofreciéndose a través de Internet, con acceso libre, a lo largo de los últimos dieciocho años. Recientemente se han integrado en el portal *BIDISO. Biblioteca Digital Siglo de Oro* <<http://www.bidiso.es>>. La colaboración de este grupo de especialistas en Humanidades con otro de especialistas en Informática en la misma universidad (el LBD: Laboratorio de Bases de Datos) <<http://lbd.udc.es/>> ha resultado mutuamente enriquecedor. Se presentan aquí las razones de la creación del SIELAE y el LBD, su finalidad, los logros en dos de las cuatro líneas de investigación (Emblemática y Polianteas) y los detalles técnicos y metodología en la creación y mantenimiento de BIDISO, los aspectos clave en su desarrollo, su evolución, arquitectura, tecnología e infraestructuras actuales.

Palabras clave: Bibliotecas Digitales, Emblemática, Polianteas, Humanidades Digitales.

ABSTRACT

Since 1996, the SIELAE (Seminario Interdisciplinar para el Estudio de la Literatura Áurea Española) <<http://www.bidiso.es/sielae>> (University of A Coruña, Spain) has pro-

duced digital resources (databases and digital libraries) in around four areas of research related to the sixteenth and seventeenth centuries: Emblematics, Early News Pamphlets, Early Modern Encyclopedia (Polyantheae) and Inventories of Personal Libraries. These resources were offered through the Internet, with open access to over the past eighteen years. Recently, they had been integrated into the BIDISO portal (Biblioteca Digital Siglo de Oro) <<http://www.bidiso.es>>. The collaboration of this group of specialists in Humanities with another of specialists in Computer Science in the same university (the LBD: Database Laboratory) <<http://bd.udc.es>> has proved mutually enriching. We present here the reasons of the creation of the SIELAE and LBD, purpose, achievements in two of the four investigation lines (Emblematics and Polyantheae) and the technical details and methodology in the creation and maintenance of BIDISO, the key aspects in the development, evolution, architecture, technology and current infrastructures.

Keywords: Digital Libraries, Emblematics, Polyantheae, Digital Humanities.

Las redes sociales como corpus de estudio para el Análisis del discurso mediado por ordenador

ANA MANCERA RUEDA
Universidad de Sevilla
anamancera@us.es

ANA PANO ALAMÁN
Università di Bologna
ana.pano@unibo.it

1. INTRODUCCIÓN

El *Análisis del discurso mediado por ordenador* (ADMO) es el acercamiento a la comunicación en la Red desde la metodología del Análisis del discurso. Su principal objetivo es analizar las propiedades lingüísticas y pragmáticas de este tipo de interacciones. Uno de los aspectos que más atención ha recibido por parte de los analistas del discurso mediado por ordenador es la naturaleza híbrida, entre oral y escrita, de los textos que se publican en el entorno digital. De ahí que en el ámbito de la lengua española Francisco Yus (2001) y Leonardo Gómez Torrego (2001), entre otros, hablen, por ejemplo, de *textos escritos oralizados* para hacer alusión a algunos de los que se difunden en la Red. Desde esta perspectiva, Ana Pano (2008) explora cómo los condicionantes situacionales pueden explicar por qué los hablantes utilizan una modalidad coloquial en contextos donde lo que prima es la inmediatez comunicativa y, más recientemente, Ana Mancera (2011) afronta el estudio de este aspecto en el periodismo digital.

Sin embargo, en la actualidad, donde de manera más recurrente se hace uso de la modalidad coloquial es en las redes sociales virtuales, ámbito en el que se centran nuestros trabajos más recientes, cuyo objeto es verificar cuáles son los rasgos coloquializadores más frecuentes en los mensajes que se publican en estos canales. Para ello, en primer lugar, hemos constituido un corpus de textos publicados en Facebook, Tuenti y Twitter considerando algunas cuestiones metodológicas relacionadas con la recopilación de datos extraídos de Internet para un análisis de tipo lingüístico-discursivo (Susan C. Herring, 2001; Jannis Androutsopoulos y Michael Beißwenger, 2008). Algunas de estas cuestiones son la de la representatividad –que nos ha obligado a recopilar un número similar de mensajes en las tres redes sociales–, la cantidad de información contextual necesaria, y la privacidad –cuestión ética que tiene que ver con el anonimato de los usuarios de estas redes–. Vamos a examinarlas con más detenimiento.

2. LA LINGÜÍSTICA DE CORPUS Y EL ANÁLISIS DEL DISCURSO MEDIADO POR ORDENADOR

En los últimos años, con el desarrollo de la Lingüística de corpus (LC), se ha impuesto una tendencia a recopilar textos naturales y completos, tratando de alcanzar una relativa extensión y diversidad. La hoy denominada LC es un área de la lingüística que ha adquirido un espacio autónomo, y que se define como "una metodología para la investigación de las lenguas y el lenguaje, que permite llevar a cabo investigaciones empíricas en contextos auténticos, y que se constituye en torno a ciertos principios reguladores" (Giovanni Parodi, 2010: 15)¹. Por su parte, John Sinclair (1991: 171) sostiene que un corpus es "una colección de textos de ocurrencias de lenguaje natural, escogidos para caracterizar un estado o una variedad de lengua". La LC es una técnica, cuyo fundamento es el corpus mismo. En este sentido, y de acuerdo con distintos estudios (Geoffrey Leech, 1992; Francisco Marcos Marín, 1994; Josse de Kock, 2001; Julia Lavid, 2005, entre otros), este no sería un campo ni un área de estudio, sino un ámbito de investigación focalizado en los corpus con base en metodologías diferentes, por ejemplo, la del Análisis del discurso (Paul Baker, 2006), una cuestión que también concierne a las Humanidades digitales entendidas bien como disciplina, bien como método de trabajo en el ámbito de diversas disciplinas. Entre las ventajas de utilizar corpus se han señalado las siguientes: permiten una adecuada representación del discurso en muestras amplias y representativas de textos originales; posibilitan la exploración de textos etiquetados y no etiquetados; por medio del procesamiento (semi)automático de los textos, es posible realizar análisis más amplios y detallados; ofrece mayor fiabilidad en los análisis cuantitativos y cualitativos; los resultados son acumulativos y confrontables con posteriores investigaciones. Un aspecto relevante que persiguen hoy en día los trabajos desde la LC radica en el interés por la variabilidad lingüística. Por ello, existe una tendencia a realizar investigaciones multirregistros y/o multigéneros, en las cuales es posible establecer comparaciones entre variedades distintas de una misma lengua.

Esto es más evidente, si cabe, cuando se utilizan las páginas de la web actual para constituir corpus de estudio lingüístico y discursivo (Mike Rundell, 2000; Philip Resnik y Noah Smith, 2003; William H. Fletcher, 2004). Como señalan Adam Kilgarriff y Gregory Grefenstette (2003: 33),

"language scientists and technologists are increasingly turning to the Web as a source of language data, because it is so big, because it is the only available source for the type of language in which they are interested, or simply because it is free and instantly available".

De hecho, los principales estudios dedicados a la metodología basada en corpus aplicada al Análisis del discurso mediado por ordenador (Michael Beißwenger

¹ En el ámbito de la lengua española, destacan los corpus: Proyecto de Estudio coordinado de la norma lingüística culta de las principales ciudades de España e Ibero América; PRESEEA (Proyecto para el Estudio Sociolingüístico del Español de España y de América); CREA (Corpus de Referencia del Español Actual) y CORDE (Corpus Diacrónico del Español), de la Real Academia Española de la lengua, que prepara ahora el CORPES XXI (Corpus del español del Siglo XXI); Val.Es.Co, centrado en la lengua oral, el registro coloquial y la variedad conversacional de la lengua; COVJA, de la Universidad de Alicante y COLA, de la Universidad de Bergen, sobre las variedades juveniles del español.

y Angelika Storrer, 2008; Brian King, 2009) coinciden en que la constitución de un corpus para investigar el lenguaje en la comunicación mediada por ordenador (CMO) es una tarea aparentemente sencilla, ya que los datos están digitalizados y son accesibles fácilmente en la Red. Sin embargo, en seguida surgen incompatibilidades entre algunos parámetros de la LC y los datos recogidos en los distintos canales. Y es que los datos extraídos a partir de la web presentan un elevado grado de heterogeneidad, lo cual dificulta la obtención de un conjunto de documentos de aspecto homogéneo, estructurado y libre de "ruido"² (Fernando Martínez Santiago *et al.*, 2001). Por tanto, además de tener en cuenta aspectos fundamentales a la hora de constituir un corpus, como la extensión del mismo, la representatividad, o la ética en la recolección de datos, Claudia Claridge (2007) y Michael Beißwenger y Angelika Storrer (2007) advierten de que para establecer corpus de estudio de la CMO es necesario distinguir entre los datos que genera el sistema –por ejemplo, las líneas de entrada y salida de un canal de chat, o la fecha de envío de un *tuit*– y los que genera el usuario, que pueden incluir texto, enlaces, emoticones e imágenes.

A la hora de recolectar datos y constituir un corpus de textos extraídos de las redes sociales virtuales Facebook, Twitter y Tuenti hemos tenido en cuenta todos estos aspectos. Pero quizás una de las cuestiones más importantes y más controvertidas³ que hemos considerado es la de la representatividad. Un corpus puede ofrecer información detallada acerca de una lengua particular, pero es imposible constituir uno que abarque toda una lengua. Por tanto, cabe buscar soluciones para establecer una proporcionalidad adecuada del corpus que conduzca a ciertas proyecciones, aun renunciando a realizar generalizaciones. Además, un corpus no es una única instancia comunicativa y tampoco cuenta con un cierre de ningún tipo, sino que presenta una organización predeterminada en torno a categorías identificables para la descripción de una variedad de lengua. Partiendo de estas consideraciones, nuestro objetivo ha sido el de constituir lo que Parodi llama un "pre-corpus" de mensajes extraídos de redes sociales virtuales, "con el fin de proponer hipótesis de trabajo y de explorar ciertas características o categorías para una posterior recolección más amplia y robusta" (Giovanni Parodi, 2010: 27).

Por otra parte, y siguiendo a Elena Tognini-Tonelli (2001), cabe distinguir entre el enfoque *corpus-based* y *corpus-driven* en función de si las categorías de análisis lingüístico están previamente determinadas y enmarcadas en una opción teórica, o si emergen del análisis y dan sustento a la construcción de una teoría. En la primera opción el corpus actúa como un método de indagación y corroboración de ideas preexistentes, así, en nuestro caso, hemos adoptado ese primer enfoque para indagar sobre las marcas de la oralidad en las tres redes mencionadas y sus implicaciones en la escritura que se emplea en ellas. Con este objetivo hemos tratado de recoger una muestra lo más representativa posible de los enunciados publicados en estos espacios. El primer paso ha sido constituir dicho corpus por medio de la recopilación de 600 mensajes publicados

² "Conjunto de términos no pertinentes o de palabras no significativas que se obtiene en una búsqueda automática a una base de datos documental" (*Glosario de Terminología Informática*, <<http://www.tugurium.com/gti>>).

³ En palabras de Adam Kilgarriff y Gregory Greffentette: "'Representativeness' begs the question 'representative of what?' Outside very narrow, specialized domains, we do not know with any precision what existing corpora might be representative of" (2003: 333).

entre noviembre de 2011 y abril de 2013 en perfiles muy diferentes, creados tanto por jóvenes como por políticos, escritores, periodistas, profesores o empresarios. Estos datos preliminares constituyen un primer panorama singular, pues se cuenta con pocos antecedentes similares –corpus de mensajes publicados en distintos canales– producto de investigaciones de naturaleza comparativa en lengua española.

En la LC, el segundo paso prevé una descripción lo más detallada posible del contexto de recopilación de los mensajes, que contribuye a justificar la selección de estas redes frente a otros canales de comunicación en la Red, y a identificar las principales semejanzas y diferencias entre los tres espacios, parámetros esenciales de cara al análisis comparativo. Para ello hemos consultado el informe *5ª Oleada del Observatorio de redes sociales* (abril, 2013). En este documento se señala que Facebook, Twitter y Tuenti son las redes más utilizadas y con el mayor número de usuarios en España. La primera es líder, con un 83% de los entrevistados que tiene una cuenta personal y la utiliza, frente a Twitter (42%) y Tuenti (27%). Los motivos principales que llevan a utilizar estas redes son, en el caso de Facebook y Tuenti, reforzar el contacto con el círculo social cercano, actualizando el perfil y compartiendo fotografías, y en Twitter, informarse y promover o apoyar causas sociales o solidarias, además de seguir a personajes famosos. La media de edad de los usuarios es de 31 años en Facebook y de 28 en Twitter. En cambio, en Tuenti, el 59% tiene menos de 25 años. Así, las redes sociales se distinguen básicamente por su función y por el tipo de usuario, sus características y objetivos.

Para Inmaculada Berlanga y Estrella Martínez (2010), la función del lenguaje que predomina en Facebook es la conativa o apelativa, el elemento preponderante es el receptor y la intención comunicativa influir en él por medio de instrucciones, consejos y preguntas. De hecho, abundan aquí los enunciados exhortativos e interrogativos, el modo imperativo e indicativo, la segunda persona verbal y los vocativos. Decíamos que Tuenti es la red social más utilizada por adolescentes y jóvenes universitarios españoles quienes, según Alba Torrego (2010 y 2011), están ahí para tener una constante presencia virtual y para interactuar con sus amistades. Esta investigadora señala que en dicho espacio los jóvenes juegan con el lenguaje prescindiendo en cierta medida de la ortografía y de la gramática normativas. Las razones que impulsan a los jóvenes a hacer un uso no normativo de la lengua en Tuenti pueden explicarse por la comodidad o la rapidez del proceso de escritura, esto es, por la escasa planificación en la elaboración del discurso, mientras que otros fenómenos pueden obedecer a factores de naturaleza pragmática, como el deseo de llamar la atención del interlocutor o de configurar la propia identidad diferenciándose de los demás. Por último, Twitter es una plataforma de *microblogging* o *nanoblogging*, es decir, un servicio en línea que permite enviar y publicar mensajes de no más de 140 caracteres. En este canal los mensajes tienen distintas funciones comunicativas (Tíscar Lara, 2012), como las de reconocimiento, cuando se *retuitean* los textos de otros y se reconoce su autoridad sobre la información que se comparte; dialógica, que permite conversar con alguien insertando "@usuario" en el mensaje o simplemente haciendo *clik* sobre el botón *Respuesta*; discursiva, mediante la incorporación de etiquetas, facilitando el seguimiento de distintos *tuits* sobre un mismo tema; e identitaria, pues en el perfil personal del usuario aparecen fotografías y una breve descripción que le identifican.

En lo que se refiere a la privacidad, Lori Kendall (2002: 60) ha puesto de manifiesto cómo en el Análisis del discurso mediado por ordenador cabe entender dicho concepto de otro modo, quizás porque en estos espacios "talk tends to blur the distinction between public and private". Por otro lado, la cuestión de la identidad digital es muy compleja, y tiene que ver también con el carácter virtual y público del medio, y con la tendencia a reelaborar la propia identidad en las descripciones personales, por ejemplo, en clave humorística (Ana Pano y Ana Mancera, en prensa). En todo caso, sí es posible adoptar algunas soluciones relativas a la privacidad de los usuarios en la recolección de datos. Por ejemplo, hay que tener en cuenta que en el caso de Tuenti, al tratarse de una red social dirigida principalmente a adolescentes, son estos los autores y los destinatarios de la mayor parte de los mensajes. Esto ha hecho necesario proteger su privacidad de cara al análisis discursivo. Por eso hemos suprimido cualquier tipo de información que pudiera arrojar pistas sobre la identidad de los autores o los destinatarios de tales textos. Pero el caso de Twitter es diferente, ya que aquí los perfiles no suelen ser de acceso restringido y cualquier internauta puede consultar lo publicado por otros. Sin embargo, por coherencia con nuestra manera de proceder con los mensajes extraídos de Facebook y Tuenti, hemos optado por eliminar también en el caso de Twitter el nombre de la mayor parte de los *tuiteros*. Sólo mantenemos su identidad cuando se trata de personas que desempeñan una actividad pública, y cuyo conocimiento resulta relevante para interpretar los resultados del análisis.

Por otra parte, y a pesar de que somos conscientes de que la lectura de este tipo de mensajes no siempre resulta fácil, hemos optado por reproducirlos todos tal y como fueron publicados. Es decir, sin omitir los errores ortográficos ni aquellas otras muestras representativas del subcódigo escrito que está difundiéndose en Internet.

3. ANÁLISIS DEL CORPUS: ASPECTOS ORTOGRÁFICOS Y SELECCIÓN LÉXICA

La "ciber-ortografía" presenta desafíos enormes para la LC, sobre todo a la hora de etiquetar este tipo de textos, aunque empiezan ya a aparecer algunos programas especialmente diseñados para el Análisis del discurso mediado por ordenador (Brian King, 2009: 313). En su afán por recrear la lengua oral, muchos internautas redactan enunciados seseantes como el siguiente:

"negroooo felicidadeees!! De regalo otra vueltesita en el coche jajajaja" (Tuenti, 21-04-2013).

Y son también frecuentes los casos de aféresis,

"Norabuena por duplicado, Antonio..." (Facebook, 20-04-2012).

o la supresión de algunos sonidos finales e incluso de sílabas enteras:

"ya tio pero tu ere mas grande ynmas fuerte...jaja es k ere especia" (Tuenti, 18-03-2013).

De esta forma, los internautas parecen querer recrear los rasgos dialectales del español meridional, a lo que parece responder también el uso de la hache en palabras que no la llevan, tal vez para simbolizar una aspiración:

"sevillanita miaa,ya esty en mi tierra,q hartera de autobus tia,como llevas esa fiestecilla??un besazoo muak" (Tuenti, 29-09-2012).

El hecho de que la representación del sonido africado palatal sordo se lleve a cabo en lugar de con el dígrafo ch, con una equis,

"yo exo asta la encuesta y to y por eso me lo quite y se lo dixo tiaa XD" (Twitter, 22-04-2013).

o el trueque de qu por la letra k para representar al fonema oclusivo velar sordo,

"[...] yo creo que me voy a pasar, y ya que hoy estamos todos medio muertos lo propongo como *kedada* para mañana [...]" (Facebook, 09-03-2013).

e incluso la omisión de la hache

"Xurraaaa papa ma dxo k si este año kiero aser snow! Jaja le dxo k sii! Lo ases cnmgo?" (Tuenti, 25-01-2013).

son rasgos que suelen atribuirse a la economía lingüística. Sin embargo, de las redes sociales analizadas, solamente Twitter cuenta con condicionamientos técnicos que limitan la cantidad de grafemas, lo que justificaría el uso recurrente de abreviaturas y el aspecto "jibarizado" de los mensajes (José Ramón Morala, 2001). No obstante, creemos que este tipo de escritura puede interpretarse como una marca grupal, como una muestra de que el internauta está al tanto del peculiar código comunicativo vigente en las redes sociales. Por tanto, su empleo constituye también una prueba de su integración en la comunidad virtual.

Con frecuencia, las faltas de ortografía en las redes sociales suelen atribuirse también a la falta de concentración, o a la dejadez de quien escribe desde la proximidad comunicativa, consciente de que el destinatario de su mensaje –generalmente un amigo– se encuentra al tanto de las convenciones que rigen la comunicación en las redes sociales. Pero, en realidad, estos internautas hacen uso de una ortografía diferente de la de los textos convencionales, de una "antiortografía" (Gabriela Palazzo, 2005), que no impide que el enunciador y su enunciatario se comprendan, dado que comparten las mismas competencias lingüísticas. En este sentido, José Martínez de Sousa (2004) distingue entre "faltas de ortografía" y "heterografías". Las primeras son fruto de la ignorancia de las normas que rigen la grafía del español, mientras que las segundas constituyen lo que podríamos llamar desviaciones intencionadas. A este último grupo parece responder la duplicación de vocales y consonantes que encontramos en muchos mensajes:

"#ParejasQueMeGustan peter lanzani y lali esposito definitivamente, *vuelvann mierdaaa!*" (Twitter, 20-04-2013).

Tales transgresiones pueden suponer un problema para la competencia lingüística de algunos jóvenes. Por ejemplo, al aumentar su inseguridad sobre la escritura del verbo auxiliar *haber*:

"Si mo fuera x @yomismaeirene no *abria* entendido la película ocean's eleven" (Facebook, 24-03-2013).

Del análisis de estos mensajes se desprende que la mayor parte de las "acometidas" contra la ortografía del español realizadas tiene lugar en Tuenti y, en

menor medida, en Facebook. Sin embargo, no podemos decir lo mismo de los mensajes redactados en Twitter, donde parece existir un mayor respeto a las normas ortográficas. Tal vez porque sus usuarios son más conscientes de que en la mayor parte de los casos se dirigen a internautas para los que resultan desconocidos, y para los que, más que las fotografías, su carta de presentación son las palabras. Por eso, quien no sabe escribirlas con corrección se arriesga a duros ataques, como los que recibió María Antonia Trujillo –ex ministra de Vivienda del primer gobierno de José Luis Rodríguez Zapatero–, quien publicó en su perfil de Twitter el siguiente texto, con el que trataba de mostrar su indignación después de haberse hecho pública su declaración de bienes:

“Mi anti #ff para todos los q han informado tanto de nuestro patrimonio y tampoco de nuestro trabajo en #Congreso” (Twitter, 09-09-2011).

La falta de ortografía presente en este mensaje desencadenó toda una avalancha de críticas en Twitter, y además fue objeto de burla en numerosos medios de comunicación, de modo que la ex ministra no tuvo más remedio que rectificar y aclarar en otro mensaje que no había querido decir “tampoco” sino “tan poco”. Por tanto, es importante respetar la ortografía incluso en las redes sociales, especialmente si se es un personaje público.

Por otro lado, la selección léxica resulta también de considerable relevancia a la hora de redactar estos mensajes. De hecho, Sigfrid Soria, ex diputado autonómico del PP por la isla de Lanzarote, fue apartado de sus responsabilidades en este partido el pasado mes de abril a raíz de la publicación en su perfil de Twitter de una serie de textos como los siguientes, en los que amenazaba a los miembros de la Plataforma de Afectados por la Hipoteca que se manifestaban frente a los domicilios de algunos políticos:

“Son tan cobardes los tuiteros perroflautas como lo eran los del ku klux klan cuando linchaban a seres humanos con sus identidades ocultas” (Twitter, 10-04-2013).

“Eso sí, como un perroflauta me acose por la calle, me intimide o agreda, la ostia que se lleva ni se la va a creer” (Twitter, 10-04-2013).

La selección léxica aquí resulta poco adecuada para el rol público que debe desempeñar un político. Pero además es muy llamativa la falta de ortografía cometida en la redacción del sustantivo “hostia”. Y es que sorprende mucho el empleo por parte de un político de este término, en su acepción vulgar y malsonante. Con estas palabras trataba de justificar en un programa radiofónico su escritura “antiortográfica”:

“No la escribo con ‘h’ pues no quiero semejanza alguna con la de la principal acepción de la RAE”.

Desconocemos si en este caso nos encontramos ante una verdadera muestra de lo que hemos denominado *heterografía*, es decir, de un uso innovador del lenguaje. O si esta es en realidad una mera falta de ortografía que luego el político ha tratado de disimular con sus declaraciones.

También es muy frecuente que los internautas publiquen mensajes en los que utilizan expresiones disfemísticas,

“qeee diseee loco d la morciyaa!!!! q laa staba scuxaando y la e pstoo d stadoo soo mierdaa asik caya la boca!! ajajajaj” (Tuenti, 15-12-2012).

incluso adjetivos de significación peyorativa como *feo* o *cabrón* que, en virtud del contexto, adquieren un sentido ponderativo:

“Felicidades cabroooooon !!!dejate veeeeeeeer feooo y pasatelo flamaaa jajajaja”
(Tuenti, 06-04-2013).

Tal uso resulta muy habitual en esta red social, lo que podría explicarse por dos razones: por una parte, porque Tuenti cuenta con mayor número de perfiles de adolescentes, y para estos tales prácticas comunicativas son muy recurrentes; por otra parte, porque, además, en esta red social el acceso es restringido; es decir, sólo pueden acceder los “amigos”, esto es, personas que aceptan una invitación por parte del usuario a formar parte de su red de amistades en este espacio. Por el contrario, quizás el carácter menos restrictivo de Facebook y Twitter, así como la pluralidad de edades y profesiones de sus usuarios expliquen el hecho de que hayamos encontrado un menor número de insultos y expresiones disfemísticas en estas dos redes sociales.

4. CONCLUSIONES

A través de esta rápida panorámica, esperamos haber mostrado cómo los textos extraídos de las redes sociales virtuales para llevar a cabo análisis del discurso mediado por ordenador en lengua española constituyen un interesante corpus de datos lingüísticos y de discursos heterogéneos que presentan una gran variedad de registros y de estilos. Un aspecto que plantea numerosos desafíos para el Análisis del discurso y sus métodos de investigación, en el contexto de las Humanidades digitales.

Asumiendo los principales postulados de la Lingüística de corpus relacionados con el método de recolección de datos, la representatividad de los mismos y la privacidad, en este trabajo hemos presentado un “pre-corpus” constituido por textos extraídos de tres redes sociales distintas: Facebook, Twitter y Tuenti. Este conjunto de textos nos ha permitido explorar, a través de un análisis cualitativo comparativo, las semejanzas y diferencias relativas a la ortografía y a la selección léxica en el español utilizado en las tres redes sociales, y establecer hipótesis de trabajo para futuras investigaciones centradas en las posibles causas del carácter antiortográfico y heterográfico de la escritura digital, o en los distintos grados de coloquialización de la lengua en estos espacios.

Desde un punto de vista metodológico, en próximos análisis cabrá considerar la cuestión de la anotación lingüística con etiquetadores morfológicos (*tagger*) y sintácticos (*parser*), de cara a afinar el análisis lingüístico y poder tratar los datos de modo automatizado para obtener resultados no sólo cualitativos sino también cuantitativos –frecuencia de uso de determinadas palabras o comparación estadística de los fenómenos estudiados–. De hecho, para la anotación y marcaje estructural de textos extraídos de las redes sociales, contamos ya con un modelo de etiquetado de discursos mediados por ordenador propuesto por Michael Beißwenger *et al.* (2012) y basado en los presupuestos de la Text Encoding Initiative, que puede resultar de gran utilidad para nuestros propósitos.

Por último, cabe decir que la manera de entender un corpus ha evolucionado y que, sobre todo en el caso de textos digitales publicados en la web, la explotación del mismo obliga a llevar a cabo conceptualizaciones y análisis multi-

dimensionales. Por ejemplo, en nuestro caso es necesario reflexionar en torno a aspectos como el contexto de publicación de los mensajes, pues estos se insertan en entornos semióticos complejos constituidos por fotografías, videos, etiquetas o enlaces que están estrechamente relacionados con los enunciados, y que en muchos casos determinan el sentido de los mismos.

BIBLIOGRAFÍA

- Androutsopoulos, Jannis y Beißwenger, Michael, "Introduction. Data and Methods in Computer-Mediated Discourse Analysis", *Language@Internet*, 5 (9) (2008), <<http://www.languageatinternet.org/articles/2008/1609/introduction.pdf>> [21/09/2013].
- Baker, Paul, *Using Corpora in Discourse Analysis*, London, Continuum, 2006.
- Beißwenger, Michael y Storrer, Angelika, "Corpora of computer-mediated communication", en A. Lüdeling y M. Kytö (eds.), *Corpus linguistics. An international handbook*, vol. 1, Berlin/New York, Mouton de Gruyter, 2008, pp. 292-308.
- Beißwenger, Michael et al., "A TEI Schema for the Representation of Computer-mediated Communication", *Journal of the Text Encoding Initiative*, 3 (2012), <<http://jtei.revues.org/476>> [21/09/2013].
- Berlanga Fernández, Inmaculada y Martínez, Estrella, "Ciberlenguaje y principios de retórica clásica. Redes sociales: el caso Facebook", *Enl@ce: Revista Venezolana de Información, Tecnología y Conocimiento*, 7 (2) (2010), pp. 47-61.
- Claridge, Claudia, "Constructing a Corpus from the Web: Message Boards", en M. Hundt, N. Nesselhauf y C. Biewer (eds.), *Corpus Linguistics and the Web*, Amsterdam, Rodopi, 2007, pp. 87-108.
- Fletcher, William H., "Facilitating the compilation and dissemination of *ad hoc* web corpora", en G. Aston, S. Bernardini y D. Stewart (eds.), *Corpora and Language Learners*, Amsterdam, John Benjamins, 2004, pp. 273-300.
- Gómez Torrego, Leonardo, "La gramática en Internet", *Lengua y escritura en Internet: Tres décadas de "red-acción"*, 2001, <http://congresosdelalengua.es/valladolid/ponencias/nuevas_fronteras_del_espanol/4_lengua_y_escritura/gomez_1.htm> [21/09/2013].
- Herring, Susan C., "Computer-mediated discourse", en D. Schiffrin, D. Tannen y H.E. Hamilton (eds.), *Handbook of Discourse Analysis*, Oxford, Blackwell, 2001, pp. 612-634.
- Kendall, Lori, *Hanging out in the virtual pub: Masculinities and relationships online*, Berkeley, University of California Press, 2002.
- Kilgarriff, Adam y Grefenstette, Gregory, "Introduction to the special issue on the web as corpus", *Computational Linguistics*, 29 (3) (2003), pp. 333-348.
- King, Brian, "Building and analyzing corpora of Computer-mediated communication", en P. Baker (ed.), *Contemporary Corpus Linguistics*, London, Continuum, 2009, pp. 301-320.
- Kock, Josse de, *Lingüística con corpus. Catorce aplicaciones sobre el español*, Salamanca, Universidad de Salamanca, 2001.
- Lara, Tíscar, "Twitter y sus funciones comunicativas", *Blog Tíscar.com*, 2012, <<http://tiscar.com/2012/03/11/twitter-y-sus-funciones-comunicativas>> [21/09/2013].
- Lavid, Julia, *Lenguaje y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*, Madrid, Cátedra, 2005.
- Leech, Geoffrey, "Corpora theories of linguistic performance", en J. Svartvik (ed.), *Directions in Corpus Linguistics*, Berlin/New York, Mouton de Gruyter, 1992, pp. 105-122.
- Mancera Rueda, Ana, *¿Cómo se "habla" en los cibermedios? El español coloquial en el periodismo digital*, Bern, Peter Lang, 2011.
- Marcos Marín, Francisco, *Informática y Humanidades*, Madrid, Gredos, 1994.
- Martínez de Sousa, José, *Ortografía y ortotipografía del español actual*, Gijón, Trea, 2004.

- Martínez Santiago, Fernando *et al.*, "WWW como Fuente de Recursos Lingüísticos para su Uso en PLN", *Procesamiento del lenguaje natural*, 27 (2001), pp. 1-7.
- Morala Rodríguez, José Ramón, "Entre arrobas, eñes y emoticonos", *II Congreso Internacional de la Lengua Española. El español en la Sociedad de la información*, Valladolid (2001), <http://congresosdelalengua.es/valladolid/ponencias/nuevas_fronteras_del_espanol/4_lengua_y_escritura/morala_j.htm> [21/09/2013].
- Palazzo, Gabriela, "¿Son corteses los jóvenes en el chat? Estudio de estrategias de interacción en la conversación virtual", *TEXTOS de la CiberSociedad*, 5 (2005), <<http://www.cibersociedad.net>> [21/09/2013].
- Pano Alamán, Ana, *Dialogar en la Red. La lengua española en chats, e-mail, foros y blogs*, Bern, Peter Lang, 2008.
- Pano Alamán, Ana y Mancera Rueda, Ana, "Identidades y cuentas parodia en Twitter: análisis de la ironía y del humor verbal", comunicación presentada en el Congreso Internacional de Lingüística Hispánica sobre "Lenguaje e identidad en el mundo hispanohablante", Asociación Internacional para el Estudio del Español en la Sociedad (SIS/EES), Londres, 3-5 de julio, 2013.
- Parodi, Giovanni, *Lingüística de corpus: de la teoría a la empiria*, Frankfurt, Iberoamericana Vervuert, 2010.
- Resnik, Philip y Smith, Noah A., "The Web as a parallel corpus", *Computational Linguistics*, 29 (3) (2003), pp. 349-380.
- Rundell, Mike, "The biggest corpus of all", *Humanising Language Teaching*, 2 (3) (2000), <<http://www.hltmag.co.uk/may00/idea.htm>> [21/09/2013].
- Sinclair, John, *Corpus, concordance, collocation*, Oxford, Oxford University Press, 1991.
- Tognini-Tonelli, Elena, *Corpus linguistics at work*, Amsterdam, John Benjamins, 2001.
- Torrego González, Alba, "'Eskriibo en el Tuenti komo pronunoiho'. Apuntes sobre la ortografía en una red social", *Tarbiya: Revista de investigación e innovación educativa*, 41 (2010), pp. 33-51.
- Torrego González, Alba, "Algunas observaciones acerca del léxico en la red social tuenti", *Tonos digital: Revista electrónica de estudios filológicos*, 21 (2011), <<http://www.tonosdigital.es/ojs/index.php/tonos/article/view/659/470>> [21/09/2013].
- Yus Ramos, Francisco, *Ciberpragmática 2.0. Nuevos usos del lenguaje en Internet*, Barcelona, Ariel, 2001.



RESUMEN

El propósito de este artículo es demostrar cómo las interacciones que tienen lugar en las redes sociales pueden aportar multitud de datos lingüísticos y de discursos heterogéneos que presentan una gran variedad de registros y de estilos. Todo ello convierte a este tipo de textos en un corpus ideal para su estudio desde la metodología del Análisis del discurso mediado por ordenador, una disciplina cuyo principal objetivo es la investigación de las propiedades lingüísticas y pragmáticas de los productos discursivos de impronta digital. Además, en este trabajo nos hemos propuesto realizar una revisión de los postulados centrales de la Lingüística de corpus, haciendo especial hincapié en las ventajas de establecer un corpus para estudiar documentos extraídos de la Red, pero sin dejar de lado los problemas que ello puede conllevar, como la selección del método más adecuado para la recolección de datos, la necesidad de que estos sean lo suficientemente representativos, y la importancia del respeto a la privacidad de los usuarios. Teniendo en cuenta estas cuestiones hemos constituido un corpus de mensajes extraídos de las redes sociales Facebook, Twitter y Tuenti, que nos permitirán analizar cómo se manifiesta en ellos la variación lingüística y, en concreto, la modalidad de uso coloquial, por medio de una serie de recursos ortográficos y de una peculiar selección léxica. En definitiva, en esta investigación pretendemos ofrecer una panorámica que recoja los principales desafíos a

los que debe hacer frente el Análisis del discurso mediado por ordenador y sus métodos de investigación, en el contexto de las Humanidades digitales.

Palabras clave: Análisis del discurso mediado por ordenador, Lingüística de corpus, redes sociales, variación lingüística, español coloquial.

ABSTRACT

The aim of this paper is to show how the interactions that take place in social networks can provide plenty of linguistic data and heterogeneous discourses presenting a variety of registers and styles. This type of texts can constitute a useful corpus for the research, using the methodology of Computer-Mediated Discourse Analysis, a discipline whose main objective is the study of linguistic and pragmatic properties of digital discursive products. Moreover, in this study we aimed to make a review of the main theories of Corpus Linguistics, with particular emphasis on the advantages of establishing a corpus to study documents extracted from the Internet, but without neglecting the problems this can lead, such as selecting the most appropriate method for data collection, the need for these data to be sufficiently representative, and the importance of respecting the privacy of users. According to these issues we have created a corpus of messages extracted from three social networks: Facebook, Twitter and Tuenti. It will allow us to explore in them the linguistic variation and, in particular, the samples of colloquial use, by a number of resources of a peculiar spelling and lexical selection. In conclusion, in this research we want to give an overview to the most important challenges that must face the Computer-Mediated Discourse Analysis and its research methods in the context of Digital Humanities.

Keywords: Computer-Mediated Discourse Analysis, Corpus Linguistics, Social Networks, Linguistic Variation, colloquial Spanish.

La Biblioteca Digital Ovidiana y el patrimonio bibliográfico español¹

PATRICIA MEILÁN JÁCOME
Universidad de Santiago de Compostela
patricia.meilan@usc.es

FÁTIMA DíEZ PLATAS
Universidad de Santiago de Compostela
fatima.diez@usc.es

La BIBLIOTECA DIGITAL OVIDIANA² se concibe como “espacio” de recopilación y difusión patrimonial, pero con una evidente vocación científica de estudio, que pretende paliar la carencia de atención que padece el Ovidio ilustrado en nuestro país y convertirse en una herramienta de trabajo para los estudios iconográficos relacionados con la obra de Ovidio, que ha sido decisiva para la transmisión del mito del material mitológico figurado. Y de manera concomitante también aspira a poner en valor una parte de nuestro patrimonio bibliográfico, facilitando la “reunión” de los ejemplares ovidianos dispersos en las distintas bibliotecas españolas.

Estos son los dos puntos claves entorno a los cuales gira la creación de nuestro sitio web³: por un lado el de la investigación, a través de la recopilación, estudio y digitalización de las ediciones ilustradas de Ovidio, y por otro, la revalorización del patrimonio bibliográfico español, gracias a la difusión conjunta del corpus compilado.

¹ Este artículo forma parte de los resultados del proyecto de investigación “BIBLIOTECA DIGITAL OVIDIANA: Las ediciones ilustradas de Ovidio. Siglos XV al XIX (III): Las bibliotecas de Castilla y León” (HAR2011-25853) financiado por el Ministerio de Economía y Competitividad.

² Este sitio web forma parte del resultado de varios proyectos de investigación financiados por el Ministerio dentro del Programa Nacional de Investigación Fundamental, I + D + I: “BIBLIOTECA DIGITAL OVIDIANA: Las ediciones ilustradas de Ovidio. Siglos XV al XIX (I): Las bibliotecas de Galicia y Cataluña” (Código: HUM 2007-60265/ARTE), “BIBLIOTECA DIGITAL OVIDIANA: ediciones ilustradas de Ovidio, siglos XV-XIX (II): las bibliotecas de Cataluña” (HAR2010-20015) y, la última fase, “BIBLIOTECA DIGITAL OVIDIANA: Las ediciones ilustradas de Ovidio, siglos XV-XIX (III): las bibliotecas de Castilla y León” (HAR2011-25853) que estará en marcha hasta finales del 2014. Dichos proyectos se plantean consecutivamente con una finalidad, contener digitalizadas todas las ediciones ilustradas de las obras de Ovidio desde el siglo XV al XIX, que se encuentran en las bibliotecas españolas públicas y privadas.

³ Sitio web <www.ovidiuspictus.es> y <www.ovidiuspictus.eu>.

Las nuevas tecnologías e internet han potenciado la creación de este nuevo tipo de bibliotecas, virtuales y temáticamente especializadas, que permiten visualizar en conjunto objetos dispersos geográficamente. En los últimos años, el interés en torno a Ovidio y las imágenes creadas para las *Metamorfosis* se ha centrado más bien en las cuestiones relacionadas con la recepción de la tradición clásica que toman forma en diversos sitios web de universidades y bibliotecas que han desarrollado proyectos de difusión de las ediciones ilustradas que se alojan en sus fondos en formato digital. Entre ellos destacan ***Metamorphosing the 'Metamorphoses': The Ovid Project*** de la Universidad de Vermont (EE.UU.)⁴, el archivo digital de la Universidad de Marburg, ***Ovids Metamorphosen***⁵, y, de una manera especial, ***Ovid illustrated: The Renaissance Reception of Ovid in image and text***, un completo sitio de Internet dentro de la biblioteca digital de la Universidad de Virginia (EE.UU.) que contiene la digitalización de una enorme cantidad de ediciones ilustradas del poema ovidiano con comentarios, bibliografía y varios hiperenlaces con otros proyectos⁶.

Siguiendo la estela de estos proyectos de difusión, nuestra Biblioteca Digital Ovidiana se configura no sólo como un proyecto de investigación, sino también como una biblioteca virtual que reúne ejemplares de las ediciones ilustradas de Ovidio que se localizan por todo el territorio español⁷.

1. METODOLOGÍA: PROBLEMAS Y SOLUCIONES

Para arrancar con nuestra labor hemos tenido que elaborar toda una metodología de trabajo, partiendo fundamentalmente de nuestros objetivos, pero también sobre la base de otras experiencias ya en marcha, a las que ya nos hemos referido en el primer apartado, que, sin dejar de resultar interesantes en ciertos aspectos, no funcionan bajo los parámetros de la herramienta de uso y difusión que nosotras proyectamos.

Partiendo de estas premisas nuestra metodología de trabajo consiste en primer lugar en los procesos de localización, digitalización y documentación de las ediciones de las distintas bibliotecas.

El primer trabajo de localización de las ediciones de Ovidio que se encuentran en las bibliotecas de Galicia, Cataluña y Castilla y León se ha realizado a través del ***Catálogo Colectivo de Patrimonio Bibliográfico Español*** que a partir de ahora citaremos como CCPB. Se trata un catálogo *online* financiado por el Ministerio de Cultura y que, como su nombre indica, pretende configurar un catálogo general en el que se engloben los fondos patrimoniales de todas las bibliotecas españolas y que tiene su equivalencia en otros países como Francia o Italia⁸. Este catálogo, el CCPB, contiene unas fichas bibliográficas bastante completas

⁴ <www.uvm.edu/~hag/ovid>.

⁵ <<http://www.fotomarburg.de/gaeste/OvidServ/ausgaben.htm>>.

⁶ <<http://etext.lib.virginia.edu/latin/ovid/ovidillust.html>>.

⁷ Hasta el momento se han incluido las ediciones que se localizan en las bibliotecas gallegas y catalanas, en correspondencia con los dos primeros proyectos de investigación y se está llevando a cabo la localización, digitalización y documentación de las ediciones ilustradas que se hayan en las bibliotecas de Castilla y León.

⁸ En Francia se encuentra el Catalogue Collectif de France abreviado con las siglas CCFR, <<http://ccfr.bnf.fr/portailccfr/>>, mientras que en Italia han creado un catálogo conocido como *Istituto Centrale per il Catalogo Unico* o ICCU <<http://www.iccu.sbn.it/opencms/opencms/it/>>.

en algunos casos, en otras bastante someras, a través de las cuales hemos hecho una primera selección de las ediciones ilustradas; no obstante, dado que el citado catálogo es un trabajo en constante actualización, los resultados que se obtienen de su consulta nunca son exhaustivos y se hace necesario acudir a cada biblioteca a consultar los ejemplares de las ediciones.

Sin intención de infravalorar el trabajo realizado por el CCPB con extraordinaria frecuencia han aparecido ediciones y ejemplares que no estaban catalogados ni documentados en sitio alguno, lo cual reitera la importancia de llevar a cabo nuestro trabajo de campo en las bibliotecas. Los motivos de estos hallazgos pueden ser de diversa índole, pero se suelen deber a que las fichas todavía no se hayan colgado en el CCPB, o que parte de esos fondos estén todavía catalogados en ficheros manuales, puesto que hemos comprobado que no todas las bibliotecas conceden la misma importancia a su fondo antiguo. Pero fundamentalmente en la mayor parte de las ocasiones el CCPB, teniendo en cuenta la ingente tarea que se le ha encomendado todavía no ha podido llegar a todas las zonas para poder cubrir todas las bibliotecas con fondo antiguo del país.

Como usuarios del CCPB el trabajo de búsqueda a partir del Catálogo nos sirvió para darnos cuenta de las necesidades que debía cumplir nuestro sitio web en cuanto a información de carácter bibliográfico pero también y más importante en cuanto al modo de acceso a dicha información. Por un lado el superficial tratamiento de algunas fichas bibliográficas del CCPB, debido a la variedad de personas que las realizan, puesto que no siempre las hacen los miembros del Catálogo Colectivo, y la importancia que luego comprobamos tenía una buena información de carácter biblioteconómico nos llevó a crear una serie de apartados para información con los que no contábamos en un primer momento. Por otro lado la imposibilidad en el CCPB de realizar búsquedas por bibliotecas o incluso por comunidades o ciudades en un catálogo de tal volumen nos ayudo a definir los diferentes accesos a la información que deberíamos tener en nuestro sitio web y de los cuales hablaré más adelante.

También detectamos, una vez vistas y documentadas las ediciones en las bibliotecas, algunos errores en cuanto a identificación de ejemplares incompletos, lo cual nos ratifica en la necesidad y utilidad de nuestro trabajo *in situ* con las ediciones y su fichado del cual hablaré a continuación.

Volviendo a la metodología de trabajo, otra de las novedades de nuestro proyecto consiste en la creación de una ficha **bibliográfico-iconográfica**. Una vez consultadas las ediciones, las documentamos una a una recogiendo los datos en esta ficha desarrollada por los miembros del proyecto de manera específica, que contiene una información más completa y detallada, de la que nos ofrecen normalmente las fichas bibliográficas al uso. Nuestro carácter de biblioteca y nuestro objeto de trabajo, el libro, nos obligaban a realizar una ficha con todas las características bibliográficas pertinentes. Pero además, nuestra función como herramienta de investigación no quería limitarse sólo al estudio del grabado y la iconografía, sino poder llegar a un campo más amplio de investigadores relacionados con el estudio del libro, de ahí la necesidad de realizar una descripción bibliográfica completa para que pudiera servir no sólo a personas interesadas en la iconografía mitológica, sino también a biblioteconomos, filólogos, literarios, entre otros. Si a esto sumamos las carencias que existen en algunas fichas bibliográficas en el CCPB o en los propios catálogos de las bibliotecas, consideramos importante realizar una buena descripción de cada ejemplar.

A continuación expondré brevemente las características de nuestra ficha bibliográfico-iconográfica para la cual se ha incluido un ejemplo de una de nuestras ediciones que se puede consultar en nuestro sitio web (Figs. 1 y 2).

Como toda ficha bibliográfica la nuestra incluye campos de datos básicos como el título completo de la edición y su colofón, el lugar y la fecha de publicación y la descripción física de la misma. Se han añadido además otros campos para reseñar una información suplementaria sobre las ediciones y los ejemplares, crucial para nuestro trabajo, como es el impresor, con el enlace correspondiente a cada impresor del *Consortium of European Research Libraries*⁹, el comitente y el ilustrador e incluye también un desarrollo pormenorizado de todo el contenido de cada edición.

No obstante, las aportaciones más novedosas de esta ficha bibliográfico-iconográfica son los apartados denominados **características de la edición** y **lista de ilustraciones**. El primero de ellos, en el que reseñamos la historia y las peculiaridades de las distintas ediciones ilustradas de la obra de Ovidio, se divide en cuatro secciones: texto y traducción, comentarios, ilustración y estructura y disposición de la página. Mediante estos cuatro apartados se intenta abarcar todo el conocimiento posible de cada edición, desde información de tipo filológico (edición del texto, traducción y comentarios), hasta otras de carácter más biblioteconómico como la organización general de la página, tipo de letra, disposición de las ilustraciones, llegando hasta el apartado que más nos interesa, el de la información del aparato gráfico mismo, donde plasmamos nuestro conocimiento obtenido del estudio de la iconografía de las ediciones ilustradas, la proveniencia de las ilustraciones, los modelos iconográficos, las ediciones príncipe de los grabados, sus reutilizaciones, etcétera.

En cuanto al apartado **lista de ilustraciones** en él se refleja el aparato de ilustraciones de cada edición completo, reseñando no sólo la ubicación de cada una ellas, sino sobre todo haciendo una lectura del episodio o episodios que se representan.

Todo este tipo de información es un trabajo que nosotros realizamos de manera única y exclusiva, puesto que no se encuentra en fichas bibliográficas ni siquiera en trabajos específicos ya que nunca se da una información tan completa y rigurosa de las ediciones. En parte se podría interpretar nuestra ficha bibliográfico-iconográfica como una "autopsia" completa de cada ejemplar.

Se incluyen, además de la biblioteca y la signatura, un listado de ejemplares de esa misma edición en las bibliotecas españolas, que no sólo proviene de la búsqueda en el CCPB sino que aumenta a medida que vamos ampliando nuestro trabajo de campo, y un apartado para la bibliografía específica de esa edición.

Cada ficha aparece plasmada en la página web con toda la información para que el internauta tenga acceso a un conocimiento mayor de la edición y del ejemplar que está viendo, información que ya no es sólo de carácter bibliográfico, sino que incluye toda una investigación en relación a la iconografía y el grabado y que consigue darle una mayor relevancia y valor a cada ejemplar conservado en las diferentes bibliotecas recogidas.

Dentro del sitio web la ficha pierde su carácter de tabla y se convierte en algo mucho más dinámico (Fig. 3). A pesar de la gran cantidad de información que

⁹ <<http://www.cerl.org>>.

Metamorfois, Viana, Lorenzo Valladares, 1589 M. Córdoba, Vol. 1589 M. BCS, Vol. 1589		Metamorfois, Viana, Lorenzo Valladares, 1589 M. Córdoba, Vol. 1589 M. BCS, Vol. 1589	
Título	[LAS TRANSFORMACIONES de Ovidio: Traducción del verso Latino, en tercetos, y algunas rimas, Por el Licenciado Viana. En lengua vulgar Castellana. CON EL COMENTO, Y EXPLICACION de las Fabelas: morales, o Philosophicas natural, y moral, y Astrologicas. DEDICADADO, LO VNDO, Y LO OTRO a Hernando de Vega Cota y Fonseca, Presidente del Consejo de las Indias. Impreso en Valladolid, por Diego Fernandez de Cordova, Impressor del Rey nuestro señor, Año, M.D.LXXXIX. CON PRIVILEGIO]		la traducción de 1581 (Venecia, Cobbi). Las quince xilografías son obra de un ilustrador anónimo, que elige para sus estampas un solo episodio para cada uno de los libros. Los episodios escogidos parecen haber sido seleccionados, con alguna excepción, de acuerdo con el criterio de representar la escena descrita en los primeros versos de cada libro. En esta edición el grabado que ilustra la historia de Atalanta, que se narra en el libro sexto, está colocado por error al principio del libro noveno, y, por otro, el grabado con el combate de Hércules y Aquiles, que se narra en el libro noveno, va ilustrando el libro sexto. Para otra versión digital de los grabados: http://www.usc.es/vidio/
Obra	Metamorfois		
Lugar de publicación	Valladolid (Valladolid)		
Impresor	Fernández de Córdoba, Diego http://hemerotecas.usc.es/tema/0000036326		
Comitente			
Ilustrador	Anónimo		
Fecha	1589		
Características de la edición	<p>Texto traducción y Traducción castellana en tercetos y octavas reales de Pedro Sánchez de Viana. Siguiendo de la tradición italiana iniciada por Anguillera, está considerada como la mejor traducción al castellano de la época por su estilo y su fidelidad al texto de Ovidio.</p> <p>Comentarios Incluye como anexo al texto las <i>Anotaciones sobre los quince libros de las Transformaciones de Ovidio</i>, del mismo Sánchez de Viana, que consiste en un extenso conjunto de comentarios sobre las fábulas y las historias mitológicas.</p> <p>Ilustración Quince xilografías, una por libro, armados en orlas con dos modelos de decoración. Desde el punto de vista de la ilustración, la edición se imbrica en la tradición italiana que hasta acompañar el comienzo de cada libro uno de los quince libros con una xilografía. Concretamente el juego de grabados de esta edición es idéntico al que Bevez la primera edición de la traducción completa al italiano de G. A.</p>		<p>El contenido de la edición se organiza en dos volúmenes. Al comienzo de cada libro se sitúa la ilustración, cuyo tamaño es un tercio de la caja del texto. Pequeñas iniciales decoradas con elementos vegetales, animales y, raramente, con figuras humanas.</p> <p>Descripción física 4º (22,3 cm.) Foliación: [20, falta 1 hoja] + 1-179 f. + [45-265, 266-310] 4 f. E. Abundantes errores de foliación propios de la edición, fallas debido a una encuadernación posterior, faltas hojas al final. Sign.: 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000.</p> <p>Contenido</p> <ul style="list-style-type: none"> — E. 91 portada; FALTA — E. 93 v. errata; FALTA — E. 92 edición del rey fechada en San Lorenzo del Escorial el 16 de abril de 1588; DESUBICADA, DESPLIEGUE DEL EPÍGRAMA. — E. 93 censura del maestro Luciano fechada el 3 de enero de 1588; — E. 93 v. sonetos hasta 95 v.; — E. 95 Epigrama al autor del licenciado Juan Jordán; — E. 93 v. carta del licenciado Pedro Sánchez

Fig. 1. Ficha bibliográfico-iconográfica.

M.BCS Vol.1589		M.BCS Vol.1589	
	de Viana y Hernando de Vega Cota y Fonseca. — E. 98 Prólogo; — E. 993 Vda de Ovidio; — E. 4952 catálogo de autores hasta E. 4983 v.;		<i>La historia de Atalanta</i> . (grabado correspondiente a la ilustración del libro 6). 11. Libro 10 (E. 104v). Odisia con un 12. Libro 11 (E. 116v). <i>La muerte de Dafnis</i> . 13. Libro 12 (E. 128v). <i>El prodigio de Andale</i> . 14. Libro 13 (E. 138v). <i>El juicio de las armas</i> . 15. Libro 14 (E. 152v). <i>Glauco y Ciceo</i> . 16. Libro 15 (E. 165v). <i>La curación de Nereo</i> .
	— E. 11 libro 2; — E. 22 v. libro 3; — E. 32 v. libro 4; — E. 44 libro 5; — E. 54 v. libro 6; — E. 65 v. libro 7; — E. 79 libro 8; — E. 92 libro 9; — E. 104 v. libro 10; — E. 115 v. libro 11; — E. 128 libro 12; — E. 138 libro 13; — E. 152 v. libro 14; — E. 165 libro 15 hasta E. 179 r.;		
	— E. 179 v. tabla de las Transformaciones; — E. 1 portada de anotaciones sobre los quince libros de las Transformaciones de Ovidio. Con la mitología de las fábulas y otras cosas; — E. 1 v. errata; — E. 2 carta del licenciado Pedro Sánchez de Viana a Hernando de Vega Cota y Fonseca; — E. 3 prólogo hasta E. 4 v.;		Observaciones Falta una hoja al principio, por lo que carece de portada. El ejemplar ha sido reencuadernado, lo cual ha modificado la colocación de las primeras hojas. Faltan los folios 311-314 pertenecientes al final de las Anotaciones. El estado de conservación es excelente, aunque el ejemplar ha sido reencuadernado y bastante pulcritudado.
	— E. 5 comienzos las anotaciones hasta E. 314; FALTAN 4 ÚLTIMOS FOLIOS E. No ^a índice de las fábulas.		Ubicación Santiago de Compostela (A Coruña). Archivo-Biblioteca de la Catedral de Santiago
			Signatura 321
			Historial <i>En libros en el E. 92 disponible</i>
			Ejemplares bibliotecas españolas Barcelona. Biblioteca de Cataluña. Sign.: Millás, 02-4º. Existen dos ejemplares más en esta biblioteca con las signaturas Res 24-8 y Res 347-8. Burgos de China (Soria). Seminario Diocesano Santo Domingo de Guzmán. Sign.: J-1651 Burgos. Facultad de Teología del Norte de España. Sign.: XJ 175 Córdoba. Biblioteca Pública del Estado / Biblioteca Provincial. Sign.: 9-347 Cuenca. Seminario Mayor de San Julián. Sign.: 199-F-12 Ojún (Asturias). Biblioteca Pública del Estado "Jovellanos". Sign.: Res. 9º-62 Oviedo (Oviedo). Biblioteca Diocesana. Sign.: 1-6-15 Lanzona (Ospiateca). Convento de Benedictinos. Sign.: C-301-4313 León. Biblioteca Pública del Estado. Sign.: FA.8071 Madrid. Biblioteca Complutense (tres ejemplares). Sign.: B01 P.L1. Res 330; B01 P.L1. 36388; B01 P.OA 2326. Madrid. Biblioteca Nacional (dos ejemplares). R/3472; R/29651 Madrid. Fundación Universitaria Española. Sign.: L172-675 Madrid. Museo del Prado. Biblioteca. Sign.: Carr3103 R. 41803 Madrid. Real Academia Española (dos ejemplares). S. Cama. 6-B-39. V-406.
Lista de ilustraciones	1. Libro 1 (E. 1). <i>La creación del mundo</i> . 2. Libro 2 (E. 11). <i>Parent en la casa del sol</i> . 3. Libro 3 (E. 22 v). <i>Céfiro y la fundación de Rosicia</i> . 4. Libro 4 (E. 32v). <i>Los Menades</i> . 5. Libro 5 (E. 44). 6. <i>La irrupción de Fátos en el banquete de Iodas de Páramo</i> . 7. Libro 6 (E. 54v). <i>Atalanta y el Aquiles</i> (grabado correspondiente a la ilustración del libro 9). 8. Libro 7 (E. 65v). <i>Jasón y los toros</i> . 9. Libro 8 (E. 79). <i>Méno y Eúclis</i> . 10. Libro 9 (E. 92).		

Fig. 2. Ficha bibliográfico-iconográfica.



Fig. 3. Ejemplar en el sitio web <www.ovidiuspictus.es>.

contiene, con la que se corría el riesgo de resultar pesada, el usuario puede dirimir mediante una serie de pestañas que partes de la información quiere visualizar y cuales no, consiguiendo así no sobrecargar la pantalla con demasiado texto y haciendo la consulta mucho más ágil.

2. EL SITIO WEB: MODOS DE ACCESO A LA INFORMACIÓN

Por otra parte y como he comentado al principio, la novedad del sitio web consiste sobre todo en ofrecer distintos modos de acceder a la información. Tras el trabajo realizado durante estos años, se ha hecho indispensable, a ojos de los miembros del proyecto, conseguir presentar la ingente cantidad de información de una forma estructurada y ordenada, pero a la vez con distintos puntos de acceso que permitan a los investigadores buscar desde cualquier campo de interés. Mejora algunos aspectos del acceso a la información que le falta al Catalogo de Patrimonio y la organiza de manera coherente y no caótica como sucede en otros sitios web de estudios ovidianos como el ya citado proyecto *The Ovid Collection* en la Universidad de Virginia (EE.UU.), sitio web con una gran base de datos sobre las ediciones ilustradas de Ovidio y una interesante información que desgraciadamente resulta muy difícil de consultar en el caos de pantallas que se van abriendo sin ningún orden estructurado.

Para evitar esto nuestra Biblioteca Digital Ovidiana potencia fundamentalmente dos tipos iniciales de ordenación que van a reflejar esos dos objetivos principales que tiene el proyecto, el de investigación y el de patrimonio: por un lado estaría el acceso patrimonial (como fondos de las bibliotecas) y por otro el cronológico (mostrando la incidencia de las ediciones por siglos). El acceso a través de las distintas bibliotecas nos proporciona un conocimiento que a través del Catalogo Colectivo no puede darse ya que podemos saber a un golpe de vista cuantas bibliotecas contienen ejemplares, cuantos ejemplares tiene cada biblioteca, e incluso cuantas bibliotecas con ejemplares existen en cada provincia

o comunidad autónoma. Este acceso a la información organizado por bibliotecas ayuda a revalorizar el patrimonio bibliográfico al verse las bibliotecas identificadas e individualizadas y sus ejemplares presentados directamente a la vista del internauta. Nuestro sitio web se convierte así en una plataforma de exposición y difusión para estas bibliotecas.

La otra forma de aproximación a los ejemplares es la organización cronológica. Los ejemplares se agrupan así por siglos. Esto nos proporciona una visión de las ediciones ilustradas de Ovidio de cada siglo que se conservan en las bibliotecas españolas y ayuda a visualizar la evolución de la forma de ilustrar a Ovidio en cada siglo, que aunque sesgada, puesto que sólo incluirá los ejemplares de las ediciones conservadas en las bibliotecas españolas, cuanto más avanzamos en el proyecto más vamos viendo que se conforma una visión bastante completa del aparato figurativo ovidiano.

Por último habría un tercer acceso directo a la información consistente en una lista de ediciones (Fig. 4). Para esta lista los miembros del proyecto decidimos una nomenclatura específica creada por nosotros que se convirtió además en una herramienta muy útil para jerarquizar y organizar la información. Esta nomenclatura identifica cada una de las ediciones mediante un código formado por el nombre de la obra, el del comentarista principal o traductor si lo tienen, el impresor, el lugar de impresión y la fecha. Conseguimos así la máxima información de la forma más breve y clara, identificando la edición y proporcionando un acceso mucho más directo dirigido hacia un público más específico.

De este modo se puede ir accediendo a cada uno de los libros, consultar la ficha donde aparece volcada toda la información bibliográfica e iconográfica tanto de la edición como del ejemplar, y visualizar las imágenes de todos los grabados que contiene, desde aquellas ediciones brevemente ilustradas con dos o tres, hasta otras en las que la profusión alcanza cifras de hasta ciento noventa grabados.



Fig. 4. Organización por lista de ejemplares.

Por último el motor de búsqueda (Fig. 5) que se ha creado para el sitio web amplía las posibilidades de acceso a la información, gracias a los numerosos campos que abarca y la utilización de las palabras clave. La mayoría de los campos contenidos en la ficha bibliográfica tienen su correspondencia en los campos del motor de búsqueda, exceptuando aquellos en los que la información es de carácter narrativo. Las búsquedas podrán realizarse por formatos clásicos como título, obra, lugar de edición o fecha, pero también incluyen otros campos más novedosos como impresor, ilustrador o tema de ilustración. Este último es sin duda otra de las grandes novedades ya que, gracias a la utilización de palabras clave, permite averiguar en cuantas ediciones aparece un motivo iconográfico, cuales son y ver los grabados asociados en los que aparece dicha imagen, personaje o episodio.

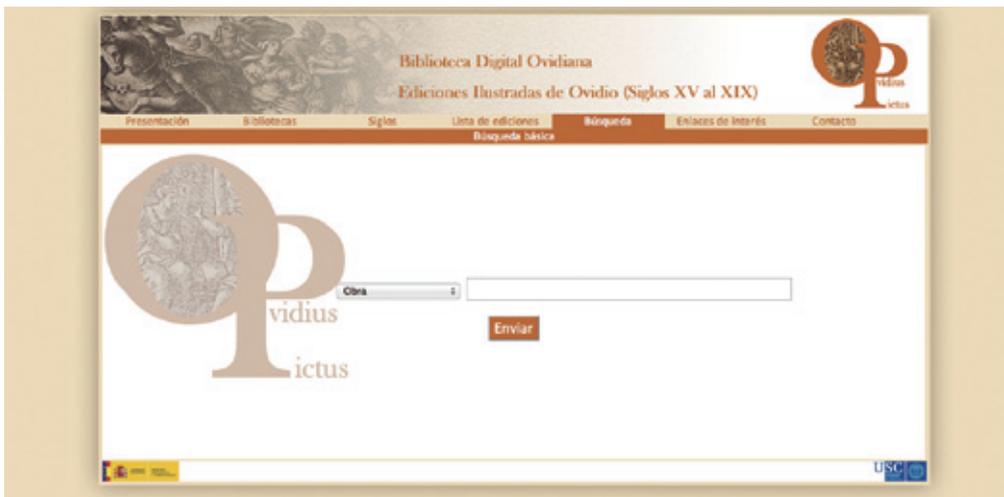


Fig. 5. Motor de búsqueda del sitio web.

3. FEEDBACK Y RESULTADOS

Además de todo esto consideramos importante la necesidad de realizar una tarea de *feedback* con las bibliotecas en las que encontramos ejemplares de Ovidio. Para ello, toda la información que producimos de sus ejemplares es enviada posteriormente en forma de esta ficha bibliográfico-iconográfica a las bibliotecas con la idea de que puedan de este modo aumentar la información que incluyen en su base de datos con respecto a estos ejemplares.

Nuestro trabajo se ve ratificado cuando en ocasiones las investigaciones llevadas a cabo corrigen errores de datación en algún ejemplar incompleto. Este sería el caso de la edición de las *Heroidas* que se encuentra en la Biblioteca Xeral de la Universidad de Santiago de Compostela¹⁰. Este ejemplar carece de las

¹⁰ *Pub. Ouidii Nasonis preclarum opus de nouo impressum: et a mendis castigatum: vna cum figuris ornatum: in quo continentur hii libri partiales: videlicet Liber heroidum epistolarum. Liber Sapphus.*

últimas hojas por lo cual no tiene colofón que nos indique la fecha de publicación que tampoco aparece referenciada en la portada. En el CCPB lo identifican como una edición de 1508 al comparar su título con otro igual que aparecía en una edición anterior de 1508 que publica el mismo librero Stephan Gueynard¹¹. Gracias a nuestro trabajo en las bibliotecas hemos podido cotejar un ejemplar de 1508¹² y constatar que esta edición sólo incluye un grabado en la portada, mientras que la edición que se encuentra en la Xeral incorpora veintiún grabados, uno por cada heroída y otro más para las obras de *Ibis*, siendo la primera edición de las *Heroidas* ilustrada impresa en Lyon 1513 de la cual se conservan muy pocos ejemplares en el mundo¹³.

Un ejemplo diferente sería la identificación de una edición de las *Metamorfosis* que se conserva en la biblioteca del Monasterio de San Francisco de Santiago de Compostela, cuyo ejemplar, en muy mal estado de conservación, carecía de portada y colofón y, por tanto, estaba sin identificar. Gracias a nuestro conocimiento de una gran número de ediciones ilustradas de Ovidio, comparando este ejemplar con otras ediciones similares llegamos a una identificación total del ejemplar como una edición lyonesa de 1528¹⁴.

Este tipo de ejemplos demuestran lo que un conocimiento experto puede aportar a los catálogos de las bibliotecas, puesto que gracias a nuestro estudio sobre los grabados de las ediciones al ver una ilustración podemos hacer una datación aproximada, a que tradición pertenece si italiana o francesa y acotando la búsqueda y comparándolo con todas las ediciones que hemos estudiado poder llegar a una identificación del ejemplar.

Para concluir este sitio web, concebido y diseñado de manera específica para contener la información sobre las obras ilustradas del poeta latino, aspira a contener un gran corpus de información sobre la obra ovidiana impresa que se encuentra en las bibliotecas españolas, de manera que se convierta, por un lado, en un medio privilegiado de difusión del patrimonio bibliográfico español y, por otro, en un valioso instrumento para el estudio tanto de la transmisión de la obra del poeta romano como de la ilustración de las ediciones impresas y la figuración de asuntos mitológicos y de otra índole. Gracias a nuestra biblioteca Digital Ovidiana pequeñas bibliotecas que no cuentan con los medios necesarios para tener un catálogo *online* o no están incluidas todavía en el Catálogo Colectivo pueden así conseguir un hueco en la red y todas ellas pueden ver parte

Libellus in Ibin. Cum expositione familiari Antonii Volsci. Ubertini clerici Cresentinatis. Domitii Calderini et Jodoci Badii singularium interpretum. Uenundantur Lugduni a Stephano Gueynard: al's Pineti Prope sanct ũ Antoni ũ. Biblioteca Xeral de la USC, signatura 13822.

¹¹ Dicha edición aparece referenciada en Baudrier (1946: 217), al igual que la edición de 1513 (Baudrier, 1946: 231), ambas con el mismo título en la portada. En ninguna de las dos Baudrier comenta que se incluyeran grabados en el texto, de ahí la posible confusión si no se han visto estas ediciones.

¹² Un ejemplar de la edición de 1508 se encuentra en la Biblioteca Reserva de la Universidad de Barcelona, signatura 07 CM-3101.

¹³ Sólo hemos localizado tres ejemplares en las bibliotecas de Carré d'Art en Nimes, en la biblioteca Louis Aragon de Le Mans y en la Universidad de Gotinga.

¹⁴ Publii Ouidii Nasonis Sulmonensis Metamorphoseos Librorum XV: opus auctum & recognitu[m].Uniuersi operis Elenchus, Ipsius Ouidii vita ex eius operibus luculenter collecta. Elegantissime Raphaelis Regii enarrationes[...] 1528 Venundatur Lugduni in vico mercuriale apud probum virum Gullielmum Boulle. Biblioteca del convento de San Francisco en Santiago de Compostela. Signatura 34-2-23.

de sus fondos dados a conocer desde una visión además totalmente experta en la materia.

BIBLIOGRAFÍA

- Adams, H.M., *Catalogue of books printed on the continent of Europe, 1501-1600 in Cambridge Libraries*, vol. II, Cambridge, 1967.
- Baudrier, H., *Bibliographie lyonnaise: recherches sur les imprimeurs, libraires, relieurs et fondateurs de lettres de Lyon au XVe siècle*, Génova, 1999 [1ª ed. 1895-1921].
- Blattner, E., *Holzschnittfolgen zu den „Metamorphosen“ des Ovid, Venedig 1497 und Mainz 1545*, Beiträge zur Kunstwissenschaft 72, Munich, 1998.
- Díez Platas, F., "Tres maneras de ilustrar a Ovidio: una aproximación al estudio iconográfico de las *Metamorfosis* figuradas del XVI", en Mª.C. Folgar de la Calle, A.E. Goy Diz y J.M. López Vázquez (eds.), *Memoria Artis. Studia in memoriam Mª Dolores Vila Jato*, tomo I, Xunta de Galicia, Santiago de Compostela, 2003, pp. 247-267.
- Díez Platas, F. y Monterroso Montero, J.M., "Mitología para poderosos: las *Metamorfosis* de Ovidio. Tres ediciones ilustradas del siglo XVI en la Biblioteca Xeral de Santiago", en A. Vigo Trasancos (ed.), *Cultura, poder y mecenazgo SEMATA*, 10, Santiago de Compostela, 1998, pp. 451-472.
- Duplessis, G., *Essai bibliographique sur les différentes éditions des oeuvres d'Ovide ornées de planches publiées aux XV^e et XVI^e siècles*, París, 1889.
- Henkel, M.D., "Illustrierte Ausgaben von Ovids Metamorphosen im XV, XVI und XVII Jahrhundert", *Vorträge der Bibliothek Warburg (1926-1927)*, *Vorträge der Bibliothek Warburg*, 1926-1927, pp. 53-144.
- Huber-Rebenich, G.; Lütkemeyer, S. y Walter, H., *Ikonographisches Repertorium zu den Metamorphosen des Ovid. Die textbegleitende Druckgraphik*, Teil II: Sammeldarstellungen, Berlin, 2004.
- Moss, A., *Ovid in Renaissance France. A Survey of the Latin Editions of Ovid and Commentaries Printed in France before 1600*, Londres, The Warburg Institute, 1982.
- Perrins, C.W.D., *Italian Book-illustrations and Early Printing. A Catalogue of Early Italian Books in the Library of C.W.Dyson Perrins*, Londres, 1994 [1ª ed. 1914], nº 170 y 171.
- Prince D'Essling, *Les livres à figures Vénétiens de la fin du XV^e siècle et du commencement du XVI^e*, Florencia-París, 1907-1914.
- Sander, M., *Le livre à figures italien depuis 1467 jusqu'en 1530: essai de sa bibliographie et de son histoire*, Milán, 1969.
- Schevill, R., *Ovid and the Renaissance in Spain*, Berkeley, University of California Press, 1913.

SITIOS WEB

- <www.ovidiuspictus.es> [28/09/2013].
- <<http://etext.lib.virginia.edu/latin/ovid/ovidillust.html>> [28/09/2013].
- <www.uvm.edu/~hag/ovid> [28/09/2013].
- <<http://www.fotomarmburg.de/gaeste/OvidServ/ausgaben.htm>> [28/09/2013].



RESUMEN

En este artículo se expondrán los objetivos y el desarrollo de las novedades de la Biblioteca Digital Ovidiana en relación con el estudio y la difusión del patrimonio bibliográfico español, mostrando las novedades que entraña la investigación sobre las ediciones

ilustradas, el estudio de los ejemplares y el tratamiento de los datos y los materiales digitales como una propuesta nueva en la relación con la información bibliográfica del fondo antiguo custodiado en las bibliotecas españolas recogido y accesible a través del Catálogo Colectivo del Patrimonio Bibliográfico Español.

Palabras clave: Ovidio, Biblioteca Digital, edición ilustrada, patrimonio bibliográfico, proyecto de investigación.

ABSTRACT

This article discusses objectives and the development of the Ovidian Digital Library in relation to the study and dissemination of Spanish bibliographic heritage, showing the new research involved the illustrated editions, the study of the specimens and treatment data and digital materials as a new proposal in relation to the bibliographic information guarded in Spanish libraries collected and accessible via Catálogo Colectivo del Patrimonio Bibliográfico Español.

Keywords: Ovid, Digital Library, Illustrated Edition, Bibliographic Heritage, Research Project.

Literatura digital o la reinención de la lectura: desde la poética de la obra abierta hacia la hipermedia

ISA BERNARDINO MESTRE

Universidade do Algarve, Centro de Investigação em Artes e Comunicação (CIAC)
isa_mestre@hotmail.com

Al igual que en las otras artes, también la literatura se ha realizado durante siglos con los medios y soportes de su tiempo. Después de la invención de la imprenta por Gutenberg y la consecuente supremacía de la cultura del libro y del texto impreso durante al menos cuatro siglos, con la aparición de nuevos medios de comunicación –telégrafo y fotografía– empieza a cambiar el panorama y a ampliar perspectivas. El texto, que en general en el libro vale por sí mismo, empieza a mezclarse y sobre todo, a trabajar en colaboración con otros medios de comunicación. La gran difusión del periódico y su rápida afirmación en el universo cultural son un ejemplo de ello, la confirmación de que la unión entre el texto y la imagen llegó para quedarse.

Más tarde, durante el siglo XX, la aparición del cine y de la televisión confirman el inicio de la llamada “era de la imagen”, con daño evidente para el texto escrito que pierde su dominio y empieza a afirmarse de modo paralelo y complementario en una cultura que valora más los medios audiovisuales.

Sin embargo, como explica Santaella (2007: 288) “a perda de hegemonia do livro não significa que qualquer media possa substituí-lo. Se pudesse, o livro já teria desaparecido (...)”. Lo que esto implica, más bien, es su reinención, la recreación de sus formas de expresión artísticas y comunicativas con el fin de adaptarse al nuevo contexto cultural en el que opera.

“Na realidade, quando surge um novo meio e comunicação, ele não substitui o anterior ou os anteriores, mas provoca uma refuncionalização no papel cultural que era desempenhado pelos meios precedentes. Via da regra, um período inicial de impacto é seguido por uma readaptação gradativa até que um novo desenho de funções se instale” (Santaella, 2007: 288).

Con la llegada de la tecnología digital el libro se enfrenta a una readaptación que se caracteriza esencialmente por una hibridación del género. Estamos delante de una literatura que no se limita al papel, que lo extrapola, que propone nuevas formas y nuevas relaciones con el lector, manteniendo al mismo tiempo una clara relación con las prácticas literarias anteriores.

ISA BERNARDINO MESTRE

«Literatura digital o la reinención de la lectura: desde la poética de la obra abierta hacia la hipermedia»
Humanidades Digitales: desafíos, logros y perspectivas de futuro. Janus, Anexo 1 (2014), pp. 329-333

En 2008, Arlindo Machado, se refiere al concepto de arte-media, que sirve para "designar formas de expressão artística que se apropriam de recursos tecnológicos dos mídia (...) para propor alternativas qualitativas". El término abarca las experiencias artísticas que utilizan los recursos tecnológicos desarrollados recientemente para potenciar el propio arte literario. Nunca antes arte y medios de comunicación estuvieron tan intensamente relacionados y entremezclados. El resultado de esta combinación da lugar a la aparición de nuevas estéticas y formas poéticas, estructuradas a partir de la hipermedia. Como resultado de la fusión entre hipertexto y multimedia, la hipermedia presenta características de ambos, añadiéndole el factor interactividad. Este lenguaje supera el concepto de hipertexto, ya que incorpora no sólo texto, sino también la información multimedia (imágenes, sonidos, videos).

En este contexto, también el papel de los lenguajes es reconfigurado. Texto, imagen y sonido coexisten hoy en el universo digital, y sobre todo en los libros digitales interactivos, mezclándose entre sí, superponiéndose, complementándose.

Aunque sea una evolución que ocurre "en directo", ya Roland Barthes (1970: 11-12) había descrito un ideal de textos compuestos por conjuntos de palabras o imágenes unidas por múltiples uniones y caminos de una textualidad siempre en abierto, nunca terminada.

Gosciola (2003: 34) caracteriza la hipermedia como la suma de "um conjunto de linguagens como a fotográfica, a sonora, a visual, a audiovisual e a cinematográfica, diferenciando-se da multimídia por ser mais interativa e permitir o acesso não-linear (...) bem como do hipertexto, por apresentar mais conteúdos audiovisuais".

Esta suma da como resultado, de hecho, un nuevo lenguaje interactivo que funciona de modo asociativo, fragmentario y rizomático, acercándose a la propia estructura del pensamiento humano. Así pues, sobre la base de esta estructura, se propone el estudio de un nuevo tipo de literatura, probablemente surgida en el último cuarto del siglo XX, y que se ve afectada por una nueva forma de mediación comunicativa.

Formulada a partir del universo hipermedia, el lenguaje por excelencia de la cultura digital, este tipo de literatura comparte con él algunas de sus características: la hibridación de lenguajes, la no linealidad, almacenamiento en bases de datos, navegación/exploración y la interactividad. Este tipo de literatura es llamada por Aarseth (1997) como literatura ergódica. Del griego *ergon* y *hodos*, trabajo y camino, respectivamente, la palabra designa un tipo de literatura que requiere del lector algún tipo de acción/decisión que le permite trazar un camino dentro de la narrativa. También Landow (2009: 24) nos da indicios de un cambio de paradigma que refleja una convergencia entre la teoría literaria y el campo de la computación que se unen para revitalizar los conceptos de textualidad, narrativa, lector y escritor.

Aunque sólo ahora está siendo explorado, esta expansión de la literatura tiene, según Nascimento (2009: 46), su raíz en la estrecha convivencia entre medios impresos, sonoros y visuales que, en su labor de mediación, ya no funcionan de manera aislada pero pasan a complementarse en conjunto y contribuyen para la expansión del modo de contar y leer historias. Para ilustrar este cambio, Mora (2012: 12) se sirve de la imagen de la Pangea para simbolizar el modo como la literatura agrega actualmente diversos medios de comunicación y lenguajes en su proceso de construcción de significados. Según el autor español este proceso

es el resultado de una "dinámica de flujos que afecta a las prácticas culturales y literarias" en su conjunto.

En este contexto, se puede asumir la posibilidad de una "nueva" literatura, marcada por la integración de los diversos medios de comunicación (texto, imagen y sonido) en una experiencia hipermediática. Sin embargo, hay que reflejar que esta "nueva" literatura se asume como una nueva propuesta narrativa o sólo recrea las formas narrativas existentes.

Como punto de partida para esta reflexión recuperaremos los estudios de Umberto Eco sobre la poética de la obra abierta. En 1962 el autor italiano propone el concepto de apertura de la obra de arte para mostrar que, como mensaje plural donde convergen varios significados, la obra de arte es, en esencia, un producto inacabado, un mensaje abierto para ser completado por su espectador cuando entra en contacto con ella. Es decir, en lugar de disfrutar pasivamente de la obra de arte, el espectador tiene la tarea de completarla, de revivirla, de remodelarla y (re)comprenderla en el momento de la interacción, de acuerdo con su experiencia. De hecho, es la pieza que completa el rompecabezas, el árbol del que florecen nuevas lecturas y nuevos significados. Al interactuar con la obra, el espectador la mantiene viva en su esencia, le da su perspectiva individual.

Según Eco (1962), esta abertura debe ser asumida como una elección del autor, que busca fortalecer el papel del espectador y no se limita a la idea de que cualquier obra de arte proporciona una relación interpretativa. La obra se asume, entonces, como una invitación al espectador para que la complete, para que le dé su toque personal y la tome como objeto de exploración y desafío que requiere de su parte una posición activa hacia múltiples perspectivas y abierta a una variedad de interpretaciones.

Si hacemos el puente con la literatura digital interactiva no es difícil darnos cuenta de que ambas coinciden esencialmente por la identificación del lector/espectador como un elemento decisivo en la construcción de significados. También en los libros digitales interactivos el lector debe asumir una postura activa de la cual depende el desarrollo de la narrativa. Es decir, está prácticamente obligado a participar porque, si no lo hace, elimina el sentido de la historia y se queda parado en ella sin posibilidad de avanzar. Se queda entonces con dos hipótesis: o bien se convierte en participante e interactúa con los elementos que se muestran en la pantalla, abriendo las puertas para avanzar en la historia, o si opta por tomar sólo el papel de espectador, se detiene en la primera página en una actividad de mera contemplación.

Hay que tener en cuenta que la estructura del libro digital interactivo está construida para requerir una acción por parte del lector. También cabe destacar que, aunque el camino de la narrativa hipermediática sea construido por el creador, que dispone cuales son las opciones posibles, pertenece al espectador/lector la tarea de dibujar su camino dentro de la narrativa, de ofrecerle su toque personal.

Por lo tanto, en una lectura comparativa, podemos asumir que los libros digitales interactivos comparten el principio de la apertura de la obra de arte, propuesto por Eco, en el sentido de que su estructura es también una estructura plural y abierta que quiere reforzar el papel del espectador/lector que, en su interacción con la obra, la completa en su esencia.

Más allá de esta relación, los libros digitales interactivos conservan otros elementos tradicionales como la página o la característica acción de hojear un

libro. Tomemos como ejemplo el libro interactivo *The Fantastic Flying Books of Mr. Morris Lessmore*, producido por Moonboot Studios, para el iPad. Si nos fijamos en toda la aplicación nos damos cuenta de que el concepto de página se conserva y que la propia linealidad de transición entre las páginas se mantiene. Entonces nos enfrentamos a la misma forma en que se recicla de lo tradicional a lo digital, cambiando sólo el contenido de la página, en lugar de sólo contener texto, incluye ahora el texto, videos, animación, sonido y elementos interactivos.

Lo mismo sucede con la acción de hojear que se mantiene intacta en un intento de imitar la acción que hemos tomado para el libro físico. Tengamos en cuenta que el propio sonido de pasar las páginas se conserva en la versión digital.

Una vez más es importante recuperar la idea de Santaella de que, de manera similar a lo que sucede con los medios de comunicación, tampoco en este caso podemos hablar de una ruptura definitiva entre el libro tradicional y el libro digital interactivo, pero en una refuncionalización en la cual coexisten elementos de ambos.

Johnson (1982) utiliza el concepto forma-cárcel para indicar que en el caso de la recreación, la autonomía total sobre el texto de origen es imposible y que, por lo tanto, el texto de destino siempre actúa como su rehén. Este mismo concepto se puede aplicar al binomio libro digital-libro digital interactivo en el sentido de que el libro tradicional actuará como forma-cárcel del libro digital interactivo, añadiéndole algunas potencialidades, pero manteniendo algunas de las principales características de su forma, independientemente del medio que lo aloja.

En 1990, Calvino dijo que "en el universo infinito de la literatura siempre se abren otros caminos para explorar, nuevos o muy viejos, estilos y formas que pueden cambiar nuestra visión del mundo." Producto de su tiempo, el libro digital interactivo es la prueba de una literatura que se reinventa a cada momento, que toma la palabra como su núcleo, pero no se limita a ella, y le agrega las otras artes en un movimiento de ruptura marcado por una cierta continuidad. Este movimiento es la prueba de que el libro comparte hoy día espacio con otros soportes sin que el concepto de literatura sea afectado. Más bien, esto aún puede ser potenciado.

Basada en una estructura asociativa, dinámica e interactiva, la literatura que se practica en los medios digitales se asume como irreverente y experimental, tratando de cuestionar el dogmatismo asociado con la literatura tradicional. Compartiendo los mismos ideales del movimiento de la poesía concreta, esta "nueva" literatura busca la exploración del lenguaje al límite, aprovechándose del visualismo del texto para crear una experiencia activa que desafía la relación convencional con el objeto libro (Coelho, 2009: 40).

Así que, en conclusión, podemos afirmar que la literatura digital, en lugar de romper con los padrones tradicionales, hace una inteligente reutilización de los mismos por medio de la tecnología, promoviendo una mayor complicidad entre lector y obra y potenciando la expansión del modo de contar y leer historias.

BIBLIOGRAFÍA

- Aarseth, Aspen, "Cybertext: Perspectives on Ergodic Literature - Chapter 1", em <<http://www.hf.uib.no/cybertext/Ergodic.html>> [01/08/2013].
- Barthes, Roland, *S/Z*, Paris, Points essays, 1970.
- Calvino, Ítalo, *Porquê Ler os Clássicos?*, Lisboa, Teorema, 1990.

- Eco, Umberto, *Obra Aberta*, São Paulo, Perspectiva, 1962.
- Gosciola, Vicente, *Roteiro para as novas mídias: do game à TV interativa*, São Paulo, SENAC, 2008.
- Johnson, Randal, *Literatura e Cinema – Macunaíma: do modernismo na literatura ao cinema novo*, São Paulo, T.A Queiroz, 1982.
- Landow, George P., *Hipertexto 3.0: Teoría crítica y nuevos medios en la era de la globalización*, Barcelona, Ediciones Paidós, 2009.
- Machado, Arlindo, *Arte e Mídia*, Rio de Janeiro, Zahar, 2008.
- Mora, Vicente Luis, *El Lectoespectador*, Barcelona, Seix Barral, 2012.
- Nascimento, José Augusto de A., *Literatura infantil e cultura hipermediática: relações sócio-históricas entre suportes textuais, leitura e literatura*, São Paulo, Universidade de São Paulo, 2009.
- Santaella, Lúcia, *Linguagens líquidas na era da mobilidade*, São Paulo, Paulus, 2007.



RESUMEN

La proliferación de nuevos medios y la rapidez con la que nuevos soportes se afirman en el universo comunicativo refuerza la necesidad de una reflexión sobre los nuevos productos culturales que nos son dados a conocer en la actualidad. El libro se afirma actualmente como uno de estos soportes que, sujeto a los cambios y transformaciones que la propia literatura ha sufrido, sigue la tendencia de integración en el medio digital. Demarcándose con dinamismo e creatividad de su imagen tradicional el libro empieza a aprender a reinventarse en contexto tecnológico actual.

Hablamos de un nuevo tipo de libro, pos-ebook, híbrido, intertextual y interactivo que irrumpe en los dispositivos digitales y se caracteriza por una actividad de recreación/transcreación de la literatura tradicional en papel. Aunque sólo ahora empiece a ser explorado, este tipo de libro nos coloca una pregunta clave: la de percibir hasta qué punto se desarrolla, en realidad, como una nueva propuesta narrativa o sólo recrea/transcrea las formas narrativas existentes.

Palabras clave: Libro, literatura, literatura digital interactiva, obra abierta, hipermedia.

ABSTRACT

The proliferation of new media and the speed which new supports appear in communicative universe reveals the need for a reflection about new cultural products. The book is, nowadays, an example of this, especially due to changes and usages that literature has been suffering and their need to adapt to the digital environment. In the context of New Media, the book wins space, in a dynamic and creative way, of his traditional image, reinventing itself in the new technological context.

We are talking about a new kind of book, pos-ebook, hybrid, intertextual and highly interactive, which erupts in digital devices and on which the theory is remarkably scarce. This type of literature leaves open a central question: realize if we are dealing with a new narrative proposal or if just recreate the existing narrative proposals.

Keywords: Book, literature, interactive digital literature, open work, hypermedia.

El Catálogo y Biblioteca digital de relaciones de sucesos: bases de datos bibliográficas, textos e imágenes¹

NIEVES PENA SUEIRO
Universidade da Coruña
npena@udc.es

SANDRA ÁLVAREZ GARCÍA
Universidade da Coruña
salvarezg@udc.es

En los últimos veinte años el avance en el campo de las Humanidades Digitales ha sido espectacular. El SIELAE (Seminario Interdisciplinar para el Estudio de la Literatura Áurea Española) ha desarrollado sus investigaciones en la Universidad de A Coruña combinando los estudios humanísticos y los progresos tecnológicos, y utilizando lo digital no sólo como recurso y soporte, sino también como medio de difusión libre de resultados de investigación. Con ese lema y esa responsabilidad se creó el *Catálogo y Biblioteca Digital de Relaciones de Sucesos*, y luego las otras bibliotecas digitales alumbradas por el SIELAE, que hoy forman el portal *Biblioteca Digital Siglo de Oro*.

La *Biblioteca Digital Siglo de Oro* (BIDISO) es un portal especializado en el estudio del Siglo de Oro español que ofrece fuentes para la investigación en la Literatura, la Historia, la Historia del Arte, la Historia del libro y de la lectura y la Historia de la Comunicación de los siglos XVI y XVII. El portal da acceso a bases de datos y ediciones digitalizadas (facsimilares y/o textos transcritos) que tiene que ver con: Literatura Emblemática, Relaciones de sucesos, Polianteas e Inventarios de bibliotecas del Siglo de Oro. El portal integra cuatro bibliotecas digitales que fueron creadas sucesivamente, desde 1993, por iniciativa de Sagrario López Poza en la Universidad de A Coruña.

Una de las bibliotecas más antiguas que ahora forman este portal es la *Biblioteca digital de Relaciones de sucesos* (Lám. 1) de la que me ocuparé aquí. Precisamente por su antigüedad, es la que ha sufrido mayor evolución tanto conceptual como de diseño de interfaz; podría decirse que constituye una muestra, bastante clara, de los avances de nuestro proyecto (e incluso del desarrollo de las Humanidades Digitales en España). Esta biblioteca da acceso libre a los

¹ Este trabajo se inscribe en el proyecto de investigación *Biblioteca digital Siglo de Oro IV* (código FFI2012-3436) financiado por el Ministerio de Economía y Competitividad del Gobierno de España, en el marco del VI Plan Nacional de I + D + i 2008-2011.

BIBLIOTECA DIGITAL SIGLO DE ORO

COLECCIONES Y RECURSOS DIGITALES

EMBLEMÁTICA	RELACIONES DE SUCEOS	POLIANTEA	INVENTARIOS
			
Libros de emblemas españoles o traducidos al español de los siglos XVI-XVIII + info	Grupo de investigación sobre relaciones de sucesos (siglo XVI-XVIII) en la Península Ibérica humanística. + info	ENCICLOPEDIAS Y RECURSOS DE ERUDICIÓN Enciclopedias, repertorios de lugares comunes y misceláneas de erudición humanística. + info	INVENTARIOS Y BIBLIOTECAS DEL S. DE ORO. + info

BIBLIOGR. RELACIONES DE SUCEOS

Biblioteca Digital Siglo de Oro (BIDISO)

Es el resultado del trabajo desde 1992 del Seminario Interdisciplinar para Estudio de la Literatura Áurea Española (SIELAE), de la Universidade da Coruña (ESPAÑA) y varios proyectos de investigación subvencionados por la Xunta de Galicia, el Gobierno de España (Plan Nacional I + D, Plan Nacional I+D+i) y el Fondo Europeo de Desarrollo Regional (FEDER). Este portal ofrece, para el uso de investigadores y público interesado, fuentes para la investigación en la Literatura, la Historia, la Historia del libro y bibliotecas y la Historia del Arte de los siglos XVI y XVII. El portal da acceso a bases de datos, ediciones digitalizadas (facsimilares y de textos transcritos) que tiene que ver con: Inventarios de bibliotecas particulares o institucionales del Siglo de Oro, Emblemática, Relaciones de sucesos, Polianteas, Enciclopedias, Repertorios de lugares comunes, Mitografías y Fuentes de erudición.

Responsable: Sagrario López Poza (Catedrática de Literatura Española), Universidade da Coruña (España).

SIELAE | XUNTA DE GALICIA | FEDER | UNIVERSIDADE DA CORUÑA

© 1996 - 2013. Sagrario López Poza | INFORMACIÓN LEGAL

Lám. 1.

frutos de investigación de una de las líneas en las que trabaja el equipo: las Relaciones de sucesos.

Las Relaciones de sucesos son documentos informativos que relatan un acontecimiento (una fiesta, una guerra, un suceso extraordinario –peste, huracán, etc.–, un viaje, etc.) y que se publicaban ocasionalmente, sin periodicidad –tan pronto como se tenía noticia del suceso o cuando convenía contarlo– (Lám. 2). Estos documentos presentan gran variedad formal y editorial: fueron escritos en prosa, en verso, combinando ambas modalidades, y pueden ocupar desde unas hojas a un librito de cordel o incluso un extenso volumen. Es frecuente que se publiquen anónimos, pero hay que señalar que en esta moda informativa parti-

57. 244-7 (n.º 24)
154. XIV. 102.3

RELACION

EN QUE SE DECLARA

TODO LO SUCEDIDO EN LOS

Estados de Flandes, desde fin de Agosto,
hasta veinte de Octubre de 1624.

EL DESAFIO DE CINCO DEL
Campo del Rey, contra cinco de Mauricio, y tambien de
los quatro baxeles de Dunquerque, y como la gente de
Mauricio intentò tomar el Castillo de Amberes,
con otras cosas sucedidas.



A auise a V. merced, como el Principe de Polonia llegó a la Corte de Bruselas, donde fue recebido muy solenemente, puso en el quarto del Archiduque que Dios a ya, y se le hizieron todas las fiestas, y regalos que se podian hazer a vn Principe de España, saluo que no comio con la Serenísima Infanta, fue a ver el sitio de Bredá, adonde le salieron a recibir el Marques Espinola, y los demas señores del campo con setenta compañías de acauallo, entrando en el quartel fe hizo vna braua batería sobre la ciudad, y ella respondió con el mismo estruendo, el dia siguiente se hizo suspensión de armas de vna, y otra parte, para que el Principe sin peligro pudiesse ver todas las trincheras, y fortificaciones en que se detuuó quatro dias. Suelto a Bruselas, partio, se dize, para Roma a catorze de Octubre, yale haciendo el gisto vn Mayordomo de la señora Infanta, hasta salir de los Estados.

En 16. de Setiembre llegó el Conde Mauricio con muy grande exercito por Santa Gertrud en Berghé, a legua y media de Bredá, con muchos aparatos, y municiones de guerra, aside barcas, como carros, y otros pertrechos, y se fortificò como V. m. verá por la estampa, y descripción que va con esta del sitio de Bredá, desde adonde vn Cavallero llamado el Baron de Breute, hijo del Conde Breute Frances que murio el año de mil y seyscientos, en el desafio de veinte contra veinte de la Compañia del Baron de Grouendoch, Governador de Bolduque, embió vn trompeta al señor de la Naue, hijo del dicho Baron de Grouendoch a desafiarle sobre la muerte de su padre, y sabiendolo el Marques Espinola hizo prender al dicho señor de la Naue, porque no saliese al desafio, de que se quexò grandemente al señor Conde Juan de Nasau, y al Teniente de Cauillos, y al señor de Apremon, y el Capitan Strienhuysen. diziendo, que su prision le causaria perpetua afrenta, el Conde Juan de Nasau respondió, que no se le diese nada, que Dios daría tiempo, para verse vengado, y en diez de Octubre, fingiendo el Conde de Nasau al Marques Espinola, que queria yr a reconocer el campo del enemigo, se partio con trezientos cauillos, y los señores arriba referidos, auiedo embiado vn trompeta al dicho Breute, que saliese con otros tantos sobre el desafio que ama pedido al señor de la Naue. Venidos al campo, determinaron que el desafio fuese de cinco a cinco, como se hizo, y fueron de la parte del Rey el Conde Juan de Nasau con los otros tres arriba nombrados, y el Teniente de la compañía de Grouendoch, de nuestra parte no fue nadie herido, em pero el Cō-

de 46

ciparon humanistas y escritores de la talla de Álvaro Gómez de Castro, López de Hoyos, Bartolomé Leonardo de Argensola, Lope de Vega o Quevedo. Durante el Siglo de Oro se produjeron en gran cantidad, pues la demanda fue cada vez mayor. La situación política y económica (crisis, guerras, corrupción, etc.) fomentó la "necesidad" de estar informado, convirtiéndose así las relaciones de sucesos en los primeros "mass media". Por su finalidad informativa, formativa y de entretenimiento, suelen considerarse el antecedente de la prensa actual.

Las relaciones de sucesos constituyen una rica fuente para la investigación literaria, histórica, sociológica, cultural, etc. A pesar de esto, pocos investigadores se habían dedicado hasta el momento a trabajar con estos documentos y no había ningún catálogo general que recogiese todas las relaciones en su período de existencia, sólo unos pocos que reunían relaciones publicadas en un período determinado (Simón Díaz, 1957; Agulló, 1966, 1975) o sobre un tema concreto (Alenda, 1903; Simón Díaz, 1982). Estas razones (la rica información que aportan y la ausencia de catálogos) fueron las que nos impulsaron a realizar un catálogo de relaciones de sucesos completo, que reuniese todas las relaciones de sucesos de que tuviésemos noticia, y que ofreciese, además, su descripción bibliográfica abreviada y su localización.

El objetivo principal del *Catálogo y Biblioteca Digital de Relaciones de Sucesos*² (CBDRS) es reunir y poner a disposición del ciudadano interesado en el estudio de la cultura española del Siglo de Oro, de forma gratuita y libre, la descripción bibliográfica de todas las Relaciones de sucesos impresas entre los siglos XVI-XVIII de que se tenga noticia, completada con la identificación de ediciones y la localización de ejemplares, y la edición digital (facsimilar o texto transcrito). Ha de tenerse en cuenta que la utilidad y la flexibilidad a la hora de recuperar los datos almacenados, fue premisa fundamental en el diseño y elaboración del catálogo; en la actualidad se ofrece una información muy rica, que, sin la existencia del catálogo difícilmente podría obtenerse, bien por desconocimiento del lugar de conservación de los documentos, por encontrarse lejos, por ser cara su reproducción, o porque es necesario invertir un período largo de tiempo para llegar a conseguir esos datos resultado del análisis cruzado. Así por ejemplo, el CBDRS nos permite saber qué relaciones se publicaron en un período de tiempo en un lugar, sobre un tema y dónde se conservan, y, además, obtener reproducción de los documentos (como muestra, puede hacerse la consulta sobre las Relaciones de sucesos impresas entre 1619 y 1621 publicadas en Madrid sobre temas histórico-políticos).

Corría el año 1993 cuando comenzamos a trabajar en esta línea de investigación y a pensar en el diseño conceptual de una base de datos relacional para catalogar estos impresos, tan numerosos y, paradójicamente, tan desatendidos por los estudiosos hasta entonces. El equipo de investigación para el estudio de las Relaciones de sucesos (siglos XVI-XVIII), dirigido por la profesora Sagrario López Poza se puso en contacto con la profesora de bases de datos de la Facultad de Informática de la UDC Nieves R. Brisaboa e inició una colaboración para elaborar el diseño conceptual de la base de datos que dio origen al catálogo. En un primer momento, tras muchas reuniones e intercambio de datos e informaciones, se diseñó un proyecto de catálogo informatizado (en el programa DBASE III, pronto actualizado a DBASE IV) que sirvió para catalogar

² En adelante abreviaremos en CBDRS, que es como se conoce ya.

las Relaciones de sucesos conservadas en las bibliotecas de la provincia de A Coruña. Ha de tenerse en cuenta que entonces Internet no era todavía de uso generalizado y lo que hoy es una biblioteca digital comenzó siendo una base de datos local en DBASE IV, que se alimentaba en un pesadísimo ordenador portátil, diseñada con ayuda del equipo de informática en 1995 para registrar las relaciones de sucesos localizadas un año antes en bibliotecas de A Coruña en el proceso de elaboración de una tesis de licenciatura. Este fue el germen del nacimiento del *Laboratorio de Bases de Datos* de la Facultad de Informática, dirigido por Nieves R. Brisaboa, y de un proyecto de investigación, que financió la Xunta de Galicia, que tenía como objetivo catalogar las relaciones de sucesos en las bibliotecas de Galicia y Portugal. Posteriormente, el grupo se amplió y se hizo más multidisciplinar y conseguimos financiación en diferentes convocatorias de proyectos nacionales para mantener y ampliar el catálogo de relaciones de sucesos. Tras algunos ajustes, muchas reuniones, algunos problemas y soluciones (Pena Sueiro, 1997) el catálogo en bases de datos locales se convirtió en catálogo *online*, gracias al desarrollo de una aplicación web con tecnologías JAVA que permitió, desde cualquier navegador, consultar la base de datos relacional central. Además, esta aplicación web dispone de una sección de administración donde los investigadores autorizados pueden modificar e incorporar nuevas relaciones de sucesos al catálogo, desde cualquier lugar, con sólo disponer de un navegador web.

En la actualidad el equipo ha aumentado considerablemente de tamaño y se ha diversificado, pues se han añadido a los investigadores de las áreas iniciales (Filología española e Informática) especialistas en Bibliografía, Historia Moderna, Historia del Arte, Historia del Periodismo e Historia de la imprenta. En el equipo han participado profesores de todas las Universidades de Galicia (UDC, UVIGO, USC) y de otros lugares de España y Europa; recientemente se ha incorporado una investigadora de Chile.

Como se ha señalado, nuestro grupo trabajó desde 1994 en colaboración con el Laboratorio de Bases de Datos en la creación de una base de datos relacional especializada para catalogar estos documentos, pues no existía ningún catálogo general de Relaciones de sucesos que diese cuenta de su rica producción y localizase su lugar de conservación actual. Desde entonces abordamos no pocos desafíos y, gracias a esto, conseguimos, a mi juicio, algunos logros. Nuestro grupo puso en línea en octubre de 1996 dos páginas web enlazadas (Láms. 3 y 4): una con información sobre el grupo de investigación *Grupo de Investigación sobre Relaciones de Sucesos* (que contaba con las siguientes secciones: objetivos del proyecto, logros, miembros del equipo, proyectos subvencionados, etc.) y otra página web, *Boletín informativo de Relaciones de sucesos* (BORESU)³ con información acerca de diferentes aspectos relativos a las relaciones de sucesos: qué es una Relación, un estado de la cuestión y una bibliografía secundaria, actualizada constantemente, sobre el tema. El grupo se ocupó del mantenimiento y actualización de estas páginas durante 17 años, desde 1996 hasta junio de 2013; fue en este año en el que ambas páginas web se integraron en el portal BIDISO constituyendo una de las cuatro secciones <www.bidiso.es/Relaciones>.

³ Anualmente, además, enviamos una copia en papel del *Boletín informativo* a los interesados (en julio salió el número 17).

Grupo de investigación sobre relaciones de sucesos (siglos XVI-XVIII) en la Península Ibérica

Miembros del equipo

Objetivos y logros de la investigación

Enlaces con páginas de interés

Acceso al Catálogo y Biblioteca digital de Relaciones de sucesos (siglos XVI-XVIII) Mediante este enlace se da acceso a la base de datos bibliográfica que contiene el catálogo de relaciones de sucesos españolas impresas de los siglos XVI al XVIII conservadas en **bibliotecas de Galicia y Portugal** y en otros lugares de España y de Europa. También se ofrece, mediante la interfaz de consulta, información de otros aspectos de interés estudiados en las relaciones de sucesos. De una parte de ellas se ofrece, además, digitalización completa de los ejemplares.

Boletín informativo sobre las Relaciones de sucesos españolas en la Edad Moderna

Sociedad Internacional para el estudio de las Relaciones de sucesos (SIERS)

Esta página informa sobre el Grupo de investigación que trabaja en la Universidad de Coruña (España) sobre las Relaciones de sucesos (1500-1800) en la Península Ibérica. En ella exponemos el curso de nuestras investigaciones con el ánimo de establecer contacto con todos los investigadores que utilicen Internet que lo deseen. Su creación y mantenimiento ha sido posible, así como toda la información que se ofrece, procedente de las investigaciones del equipo (incluida la base de datos), gracias a las ayudas económicas recibidas por el grupo de la Xunta de Galicia (Secretaría Xeral de Investigación e Desenvolvemento) y del Ministerio de Educación y Ciencia (Plan Nacional I + D) para los proyectos de investigación:

- **Catálogo informatizado de las Relaciones de Sucesos (siglos XVI-XVIII) en las bibliotecas de Galicia y Portugal** (proyecto XUGA 10504A90).
- **Catálogo e biblioteca dixital de Relaciones de Sucesos en bibliotecas de Galicia e Portugal (s. XVI-XVIII)**, Subproyecto 1, Humanidades (Proyecto XUGA10401B98), coordinado con el Subproyecto 2, Informática (PGICET999X10503A).
- **Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+i), Ministerio de Educación y Ciencia y el Fondo Europeo de Desarrollo Regional (FEDER), Proyecto: Biblioteca digital Siglo de Oro I: Emblemática, Relaciones de sucesos y Misceláneas de erudición (catalogación, digitalización y difusión vía Internet)**, código BF2003-03945.
- **Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+i), Ministerio de Educación y Ciencia y el Fondo Europeo de Desarrollo Regional (FEDER), Proyecto: Biblioteca digital Siglo de Oro II: Relaciones de sucesos, Poemáticas y fuentes de erudición en la Edad Moderna (catalogación, digitalización y difusión vía Internet)**, código HUM2005-07410/PILO.
- **Ministerio de Ciencia e Innovación y Fondo Europeo de Desarrollo Regional (FEDER): Biblioteca Siglo de Oro III**, código FF12009-08113 (subprograma PILO).

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
Fondos FEDER




Lám. 3.

Boletín informativo sobre las Relaciones de sucesos españolas en la Edad Moderna

Bienvenidos al boletín informativo sobre Relaciones de sucesos españolas (BORESU). Pretendemos que esta página sirva de vínculo a cuantos investigadores trabajen en esta parcela de la cultura española de los siglos XVI al XVIII.

La información que ofrecemos se obtiene de las actividades del **Grupo de investigación SIERS**, que desarrolla desde 1994 en la **Universidad de Coruña** (España), bajo la dirección de la catedrática de Literatura Española **Sagrario López Posa**, el **proyecto RSE-CBD** (Relaciones de sucesos españolas - Catálogo y Biblioteca Digital).

El equipo de investigación ofrece información general y bibliográfica sobre Relaciones de sucesos (los enlaces del margen) y **Acceso al Catálogo y Biblioteca digital de Relaciones de Sucesos (siglos XVI-XVIII)** que contiene descripción bibliográfica, análisis, y en ocasiones digitalización de miles de Relaciones de sucesos (obras, ediciones, ejemplares) conservadas en **bibliotecas** españolas y extranjeras visitadas por miembros del equipo. Para más detalles sobre las posibilidades de consulta, véase la descripción del Proyecto RSE-CBD.

Realización: Sagrario López Posa y Sandra Mª Fernández Vales (Universidad de Coruña, España)

Departamento de Filología Española e Latina
Facultade de Filología
Campus de Cebreiro, s/n
15701 A Coruña - ESPAÑA

Tel.: (+34) 981 16 70 00 (ext.1840 y 1887)
Fax: (+34) 981 167 151
Contacte nos nosotras

La creación y mantenimiento de esta página ha sido posible gracias a las ayudas recibidas para los proyectos de investigación subvencionados por la Xunta de Galicia (XUGA 10401B98 y XUGA10401B98) y los subvencionados por el Plan Nacional I + D (Comisión Interministerial de Ciencia y Tecnología) con códigos: TI186-1390-CO2-01 y TI1399-0329-CO2-01 y por el Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica (I+D+i), Ministerio de Educación y Ciencia y el Fondo Europeo de Desarrollo Regional (FEDER) BF2003-03945, HUM2005-07410/PILO y FF12009-08113 (subprograma PILO).

Página en línea desde Octubre de 1996.
95438 accesos

Última revisión: 12 de noviembre de 2012
© 1996-2012 Sagrario López Posa

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN
Fondos FEDER







Lám. 4.

junto con las otras tres colecciones fruto de los resultados de investigación de las distintas líneas en que trabaja el grupo SIEALE.

En el año 2001 se puso en línea el catálogo –hasta entonces sólo accesible en bases de datos locales– al que se accedía mediante un enlace desde cualquiera de las dos páginas web mencionadas anteriormente: *El Boletín informativo de Relaciones de sucesos* (ya conocido como BORESU) o la página del *Grupo de investigación* (Relaciones). El *CBDRS* era entonces una base de datos y biblioteca digital que ofrecía datos bibliográficos de ediciones y ejemplares de Relaciones de sucesos localizados en bibliotecas de Galicia y Portugal.

Desde entonces el catálogo no ha dejado de incrementarse, no sólo como fruto de su alimentación con datos de más relaciones de sucesos localizadas en bibliotecas, sino también por habersele ido añadiendo nuevas funcionalidades a las aplicaciones: posibilidad de visualizar y reproducir ejemplares (a través de un enlace o de un pdf), posibilidad de exportar los resultados de una búsqueda a pdf y enviarlos a un destinatario mediante correo electrónico, elaboración de un *thesaurus* de impresores y editores, etcétera.

En la actualidad en el *CBDRS* se ha convertido en una gran base de datos que almacena y da acceso a los datos bibliográficos de: 5.476 Relaciones, 5.687 ediciones y 5.824 ejemplares localizados en más de 184 bibliotecas de la península Ibérica y norte de Italia. Pueden consultarse también más de 1.780 reproducciones facsimilares digitales y algunas ediciones transcritas en pdf.

Todos estos datos, fruto del trabajo de localización y examen de ejemplares en bibliotecas, descripción bibliográfica, digitalización, etc. realizado por muchas personas a lo largo de varios años están disponibles *on-line* de forma gratuita, y pueden visualizarse a través de dos interfaces, mediante dos opciones de búsqueda:

- a) Búsqueda sencilla (Lám. 5).
- b) Búsqueda compleja (Lám. 6).

Debe señalarse que las posibilidades de búsqueda en ambas opciones superan con mucho a las de cualquier otra biblioteca, pudiendo así recuperar ediciones por una variedad de campos infrecuente en otras interfaces como: la fecha del acontecimiento, el lugar del acontecimiento, la tipología de la relación⁴, o la modalidad de escritura (si un texto está escrito en prosa o en verso), etc. Hemos incorporado también unos “consejos para realizar búsquedas”, muy útiles para personas que habitualmente no usan las bases de datos y no saben valerse de caracteres comodines, tan importantes cuando tratamos con textos del Siglo de Oro, en los que es frecuente que una palabra se escriba de muchas maneras (Faxardo, Fajardo, etcétera).

En la actualidad contamos con un grupo internacional, con investigadores de Europa y América, realizando las tareas de investigación previas a la alimentación de la base de datos y estamos preparando la creación de una red internacional para el estudio de relaciones de sucesos con grupos que se han sumado al proyecto con ánimo de realizar un catálogo general de Relaciones de sucesos en distintas lenguas europeas (*Digital Library of Early News Papers in Europe: DILENPE*). En nuestro catálogo, y gracias a la unión de las Humanidades con los

⁴ Sobre la taxonomía establecida, ver S. López Poza y N. Pena Sueiro (2000).

Lám. 5.

nuevos avances tecnológicos en el área de las Bibliotecas Digitales, reunimos una detallada información fruto del análisis experto, datos, imágenes y textos, que ofrecemos de manera gratuita con un solo clic (algo impensable cuando comenzamos, hace 19 años).

Además, recientemente hemos contribuido, junto con otros grupos de investigación españoles, a la creación de la red *Aracne. Red de Humanidades y letras Hispánicas*⁵ que surgió de la coordinación de varios proyectos en el ámbito de las letras hispanas con el fin de rentabilizar y buscar reconocimiento de los esfuerzos y resultados.

En el futuro está previsto que se vayan incorporando las relaciones de sucesos editadas en otros idiomas (italiano, alemán, inglés, etc.) y la ampliación del catálogo con la edición digital con marcación TEI. En lo tecnológico, entre otras tareas, se está estudiando la posibilidad de implementar versiones de nuestras páginas web adaptadas a móviles y a *tablets* (e incluso en un futuro podrían desarrollarse aplicaciones nativas para ANDROID e iOS) y también la adaptación de nuestras aplicaciones para convertirlas en proveedores de datos OAI-PMH que permitan la interoperabilidad entre bibliotecas e integrar así el *CBDRS* en recolectores internacionales (Hispana-EUROPEANA, etcétera).

El *CBDRS* es un recurso gratuito, flexible, y eficaz para el estudio de diferentes aspectos de los siglos XVI y XVII. Es un ejemplo más del cambio drástico en la forma de trabajar y de investigar que hemos sufrido en los últimos años y a los que hemos tenido que adaptarnos: la transmisión del conocimiento ya no es unidireccional, los compartimentos han dejado de ser estancos y las competencias del investigador actual son muy diferentes (ahora es fundamental valorar

⁵ Para más información, remito al trabajo de Consolación Baranda y Eduardo Rodríguez en este mismo volumen.

www.bidiso.es/RelacionesSucesosBusqueda/SelectLanguage.do?language=es

Grupo de Investigación sobre Relaciones de Sucesos (S.XVI-XVIII)
Catálogo y Biblioteca Digital de Relaciones de Sucesos (siglos XVI-XVIII)

UNIVERSIDADE DA CORUÑA [Página principal](#) [Búsqueda simple](#) / [Búsqueda avanzada](#)

~ Búsqueda avanzada de Ediciones de Relaciones de Sucesos ~

Especifique las condiciones que deben cumplir las Relaciones de Sucesos utilizando el siguiente formulario.

Título: Palabra(s) o fragmento(s) de palabra(s)

Autor: Palabra(s) o fragmento(s) de palabra(s)
 Anónimo

Lugar del acontecimiento: Palabra(s) o fragmento(s) de palabra(s)

Año del acontecimiento: Entre y Valor numérico entre 1000 y 2000
 Sin año de acontecimiento

Año de edición: Entre y Valor numérico entre 1000 y 2000
 Sin año de edición

Lugar de edición: Palabra o fragmento de palabra
 Sin lugar de edición

Editor: Palabra(s) o fragmento(s) de palabra(s)

Impresor: Palabra(s) o fragmento(s) de palabra(s)
 Sin impresor

Tipología: Acontecimientos políticos y religiosos ▾

Subgénero: ▾

Modalidad del discurso: Elija una modalidad ▾

Biblioteca: Elija una biblioteca ▾

País de la biblioteca: España ▾

Número de control: Ej: BORS 0002954

Todas las ediciones Ediciones con algún ejemplar digitalizado
 Todas las ediciones Modificaciones recientes

Buscar

Consejos para realizar las búsquedas:

Lám. 6.

la capacidad de utilizar con juicio crítico los medios, de empoderarnos como ciudadanos de los medios digitales y la capacidad de leer a partir de textos discontinuos y fragmentados, etcétera)⁶.

La combinación de las disciplinas tradicionales de las ciencias humanas (historia, literatura, arte, etc.) con las herramientas proporcionadas por la informática

⁶ Ver Henry Jenkins (2008).

(codificación de datos y metadatos, visualización de datos, recuperación de información gracias al etiquetado de textos, etc.), han dado lugar a que bibliotecas digitales como esta, que continuamos mejorando y manteniendo, faciliten al investigador sus tareas: el acceso directo, rápido y gratuito, a textos y documentos desde nuestro lugar de trabajo, aunque se conserven en bibliotecas lejanas; la posibilidad de consultar diferentes documentos a la vez; la posibilidad de cruzar datos resultado del estudio, de manera automática que nos permitan saber informaciones diversas como, por ejemplo, qué temas interesaron más, qué autores, en qué ciudades se publicaron más pliegos de cordel, que impresores tuvieron más actividad, qué se leía más en el siglo XVI... todo esto es posible gracias a lo que hoy llamamos *Humanidades Digitales*.

La interacción entre las Humanidades y la Computación ha supuesto un cambio exponencial en la manera de realizar nuestra investigación requiriendo cada vez más el trabajo en equipos multidisciplinares y la creación de comunidades que permitan una renovación epistémica (también pedagógica) y que faciliten la difusión de la información, la democratización del conocimiento, la posibilidad del conocimiento expandido...

A la vista de esto, puede afirmarse que, el *Catálogo y Biblioteca Digital de relaciones de Sucesos* contribuye, desde 1994, al desarrollo de las Humanidades Digitales en España.

BIBLIOGRAFÍA

- Agulló y Cobo, Mercedes, *Relaciones de sucesos, I. Años 1477-1619*, Madrid, CSIC, 1966.
- Agulló y Cobo, Mercedes, "Relaciones de sucesos (1620-1626)", en *Homenaje a Don Agustín Millares Carló*, Gran Canaria, Universidad, 1975, I, pp. 349-380.
- Alenda y Mira, Jenaro, *Relaciones de solemnidades y fiestas públicas de España*, Madrid, Sucesores de Rivadeneyra, 1903.
- Álvarez-García, S. et al., "Desarrollo de un compresor de textos orientado a palabras basado en PPM", en *Actas de las XIV Jornadas de Ingeniería del Software y Bases de Datos (JISBD'09)*, San Sebastián (España), 2009, pp. 237-248.
- Brisaboa, Nieves R. et al., "Interfaz de consulta a una base de datos de *Relaciones de sucesos* vía Internet", en Sagrario López Poza y Nieves Peña Sueiro (eds.), *La fiesta: Actas del II Seminario de Relaciones de Sucesos*, A Coruña, 1998, pp. 29-42.
- Cunningham, Leigh, "The Librarian as Digital Humanist: The Collaborative Role of the Research Library in Digital Humanities Projects", *Faculty of information Quaterly*, vol. 2, nº 1 (2010), <<http://fiq.ischool.utoronto.ca/index.php/fiq/article/view/15409>>.
- Dexeus, Mercedes, "La digitalización de imágenes y el estudio de la imprenta manual", en Sonsoles Celestino Angulo (coord.), *De libros y bibliotecas: homenaje a Rocío Carracuel*, 1995, pp. 107-116.
- Dacos, M., "Manifiesto por unas Humanidades Digitales", *Hypotheses.org*, 26 de marzo de 2011, <<http://tcp.hypotheses.org/487>>.
- Galina-Russell, Isabel, "¿Qué son las Humanidades Digitales?", *Revista digital universitaria*, vol. 12, nº 7 (2011), <<http://www.revista.unam.mx/vol.12/num7/art68>>.
- Jenkins, Henry, *Convergencia Culture*, Paidós, 2008.
- López Poza, Sagrario, y Nieves Peña Sueiro, "Diseño de una base de datos para catalogación y estudio de *Relaciones de sucesos*", en Luis Díaz de Viana (coord.), *Palabras para el pueblo. Vol. I, Aproximación a la literatura de cordel*, Madrid, CSIC, 2000, pp. 367-379.
- López Poza, Sagrario, "Catálogo y edición digital de *relaciones de sucesos* españolas accesible en Internet", en Aurora Miguel Alonso, Gloria Carrizo Sainero e Isabel García-Monge

- (eds.), *Trabajos de la VIII reunión de la Asociación Española de Bibliografía (2003)*, Madrid, Asociación Española de Bibliografía, Biblioteca Nacional, 2004, pp. 93-102.
- Pena Sueiro, Nieves, "Un catálogo informatizado y biblioteca digital de Relaciones de sucesos (s. XVI-XVIII): Problemas y soluciones", en *III Xornadas de Arquivos. Bibliotecas e Museos de Galicia*, Ferrol, 23-25 de septiembre de 1999, Santiago de Compostela, Xunta de Galicia, 1999, pp. 529-538.
- Peñasco González, Sandra M^a, "Presentación de la nueva interfaz de consulta del *Catálogo y biblioteca digital de relaciones de sucesos (siglos XVI-XVIII)* del grupo de investigación sobre relaciones de sucesos (S. XVI-XVIII)", en Patrick Bégrand (ed.), *Representaciones de la alteridad, ideológica, religiosa, humana y espacial en las relaciones de sucesos, publicadas en España, Italia y Francia en los siglos XVI-XVIII*, Actas del V Congreso Internacional SIERS, 6-8 de septiembre de 2007, Besançon (Francia), Université de Franche-Comté, 2009, pp. 93-99.
- Simón Díaz, José, "Algunas relaciones de sucesos de los años 1540-1650", *Biblioteca Hispana*, tomo XV, nº 4 (1957), pp. 506-523.
- Simón Díaz, José, *Relaciones de actos públicos celebrados en Madrid (1541-1650)*, Madrid, CSIC, 1982.



RESUMEN

El *Catálogo y Biblioteca digital de Relaciones de sucesos*, creado en la Universidad de A Coruña en 1994, es una muestra clara de las ventajas de la interacción entre las Humanidades y las nuevas tecnologías. En este trabajo se explica la evolución que ha sufrido desde su nacimiento, en bases de datos relacionales locales, hasta la actualidad, integrado en el portal *BIDISO*, de acceso abierto y consulta gratuita en Internet. *CBDRS* ofrece al interesado en la cultura áurea una rica información, muy útil para los estudiosos de diversas disciplinas: Literatura, Arte, Historia, Ciencias de la Comunicación, Bibliografía, etcétera.

Palabras clave: *CBDRS*, Relaciones de sucesos, Humanidades digitales, Bases de datos, Siglos de Oro.

ABSTRACT

The *Catálogo y Biblioteca digital de Relaciones de sucesos*, established in the University of A Coruña in 1994, is a clear demonstration of the benefits of the interaction between the Humanities and the new technologies. This paper explains the evolution it has undergone since its birth in local relational database, to the present, open access and free consultation online. *CBDRS* provides rich information to those interested in Golden Age culture, useful to students of various disciplines: Literature, Art, History, Communication Studies, Bibliography etc.

Keywords: *CBDRS*, News Pamphlets, Digital Humanites, Data Bases, Golden Age.

Ensayo de un índice conceptual para un corpus literario de textos alineados multilingües

SANTIAGO PÉREZ ISASI
Centro de Estudos Comparatistas, Universidade de Lisboa
santiagoperez@campus.ul.pt

JOSÉ RAMÓN CARRIAZO RUIZ
Universidad Nacional de Educación a Distancia-UNED
carriazo@flog.uned.es

1. INTRODUCCIÓN: EL PORTAL ANDRÉS DE POZA Y LOS ÍNDICES

El portal *Andrés de Poza* (<http://andresdepoza.com>) es un portal especializado en la edición de textos con versiones múltiples (o “textos multiversión”) dirigido por Carmen Isasi y desarrollado, en diversas etapas, con el apoyo de la Universidad de Deusto, el Gobierno Vasco y la Fundación Jesús de Gangoiti. Los textos incluidos en el corpus pertenecen a dos categorías:

- Textos que han tenido diversas ediciones o versiones a lo largo del tiempo (principalmente, textos notariales o jurídicos, como el Fuero de Alcalá de Henares, aunque también algunos textos literarios como *La libra* de Virgilio Malvezzi, en sus ediciones de Pamplona, Nápoles y Turín).
- Textos que han sido traducidos a varias lenguas. La mayor parte de los textos son literarios (por ejemplo, *Il Romulo* de Malvezzi o la *Introducción a la vida devota* de Francisco de Sales), aunque también hay ejemplos de documentos notariales multilingües, como la Fundación de la Cofradía de Pescadores de Bermeo.

El portal Andrés de Poza ofrece ediciones digitales múltiples alineadas dinámicamente. La visualización dinámica de los textos permite que el lector pueda seleccionar las versiones que quiere cotejar, hasta un máximo de cinco, que se presentan en columnas y alineados por unidad de traducción. Se ha empleado, para ello, un alineamiento previo semi-automático, llevado a cabo con el *software* de creación de memorias de traducción *Déjà Vu*, así como un etiquetado XML adaptado a las directrices de la Text Encoding Initiative (TEI-P5).

Desde el momento de su creación, el portal Andrés de Poza ha experimentado con diversos modos de enriquecimiento del texto y de visualización, que permitan al usuario-lector obtener tanta información como sea posible sobre los



Fig. 1. Página principal del Portal Andrés de Poza.

<p>Teme d'ogn'vno il Tiranno, ed è fatale, che tema il proprio esemplo, perche dal temer tutti non si escluda, in vn certo modo, il temer anche se stesso.</p>	<p>The Tyrant is afraid of every one, and it is fatal that he fears his owne example; for in that hee fears every one, he is not exempted from fearing, in a manner, himselfe.</p>	<p>Teme de cada uno el tirano, y es fatal que tema el propio exemplo, porque del temer a todos no se excluya, en un cierto modo, el temerse también a sí mismo.</p>	<p>Vn Tyran redoute tout; aussi est il bien iuste qu'il craigne l'effet de l'exemple qu'il donne; dans cette crainte generale de toutes choses, il semble qu'en certaine façon, la crainte de soy mesme y soit comprise.</p>	<p>Een Tyran vreeset voor een yder, en het is een onvermydelijke nootwang, dat hy sijn eigen voorbeeld vreeset, door dien van alle vreesen, in eeniger voegen niet uitgesloten word, ook sich selfs te vreesen.</p>
<p>Crede assicurarsi senza sangue bastevolmente dal fato, ponendola fra le Vestali, a consacrare la Virginità a gli Dei.</p>	<p>He thinks to fence himselfe sufficiently against fate without blood, by putting her amongst the Vestals, to consecrate her virginity to the gods.</p>	<p>Cree que se asegura sin sangre bastantemente del hado, poniéndola entre las vestales, y consagrando la virginidad a los dioses.</p>	<p>Amulus se persuade qu'en exterminant son sang et sa race, il se fortifie contre la destinee: Cela le fait resondre de mettre sa niepce avec les Vestales, et de consacrer sa Virginité aux Dieux pour luy oster le moyen d'auoir ligne.</p>	<p>Hy geloofde sonder bloedstorting sich genoegsam te versekeren wegens de hooge schikking des Noodlots, met haer onder de Verstaelsche Maegden te stellen, en haren maegdom aen de Goden te doen opofferen.</p>
<p>Serue alle donne, co'Tiranni, per innocenza la debolezza. Hanno egliuo maggior difficultà, doue hanno minor resistenza. Non possono trouar in esse quel delitto, che fà lodare la crudeltà, o fingere in se stessi quel timore, che la fà temere.</p>	<p>Womens weakness serueth for innocencie with Tyrants; they meet with more difficult, where they are least withstood; they cannot finde in them those faults which makes their cruelty be praised; or faine such feare in themselves, as may serue</p>	<p>Serue a las mugeres con los tiranos la debilidad por innocencia. Tienen ellos mayor difficultad donde hallan menor resistencia. No pueden hallar en ellas aquel delito que haze alabar la crudeltad, o fingir en sí aquel temor que la fà temer.</p>	<p>L'impuissance de ce sexe passe ordinairement pour innocencie entre les Tyrans; mais les choses qu'ils redoutent le moins, sont celles qui leur font souuent plus de mal. Les femmes ne leur fournissent pas d'assez forts pretextes pour rendre leur cruauté</p>	<p>De swakhot dient de Vrouwen, in de oogen der Tyrannen, voor onnoosihet. Sy vinden meerder moeilikhof, waer sy minder tegenstand hebben. Sy konnen in haer sulke gebreken niet ontdekken, door welke hare wreedheit sou mogen</p>

Fig. 2. Textos alineados dinámicamente.

documentos editados, así como diversas aproximaciones al texto que respondan a intereses y finalidades diversas. En este momento, además de continuar ampliando el corpus de textos y versiones en las líneas anteriormente indicadas, los esfuerzos del proyecto Andrés de Poza se centran en la creación de índices de diverso tipo que faciliten al usuario el acceso al texto y a la información que contiene. Es esta una posibilidad que venimos desarrollando desde prácticamente

el inicio del proyecto Andrés de Poza. Así, hemos implementado ya, en los textos publicados en Andrés de Poza, los siguientes tipos de índices:

- *Glosarios multilingües con y sin contexto*: Los glosarios léxicos incluidos en obras como el *Romulo* de Malvezzi ofrecen un listado multilingüe de términos significativos, que pueden visualizarse aisladamente o en su contexto amplio.
- *Índices de antropónimos y topónimos* creados automáticamente a partir del etiquetado, en obras como la *Philotea* de Francisco de Sales o *La libra* de Virgilio Malvezzi; en este último caso se han introducido diversas variantes gráficas de los topónimos (“Alsacia”, “Alsace” y “Elsass”, por ejemplo) para facilitar las búsquedas. En ambos casos (en el índice de antropónimos y de topónimos), es posible acceder a una lista con todas las apariciones del término con contexto, pinchando en el término deseado en cualquiera de sus manifestaciones gráficas.
- *Índices de contenidos*: En el caso de las Cartas Puebla de Vizcaya, que fueron objeto del proyecto *Foru Ondarea* (financiado por la Diputación Foral de Bizkaia. Bizkaiku Foru Aldundia), se ofrece al usuario un índice por contenidos, un tipo de índice favorecido por la estructura jurídica común a estos documentos.
- *Índices conceptuales*: El siguiente paso en esta línea de exploración, del que se ocupa precisamente esta presentación, es la creación de índices conceptuales jerárquicos y multilingües, que respondan a la estructura de un campo semántico determinado (el del léxico político, en este caso) y que permitan el acceso a los textos a partir de los conceptos.

Será necesario, antes de ofrecer una visión del estado actual de nuestras pruebas-piloto, establecer las bases teóricas y metodológicas previas en cuanto a los conceptos de “índice conceptual” y “léxico político”, para después ejemplificar su aplicación al texto del *Romulo* de Virgilio Malvezzi.

2. ÍNDICES, GLOSARIOS, ONOMASIOLOGÍA Y WEB SEMÁNTICA

2.1. Planteamientos teóricos y metodológicos

La filología, en este nuevo mundo de las humanidades digitales, vuelve a ganar actualidad e interés, pues sus tareas principales (la edición de textos y la creación de instrumentos para su interpretación en sentido amplio) resultan al tiempo los objetivos básicos perseguidos por los creadores de los materiales virtuales basados en la palabra. Quien fue pionero en la Text Encoding Initiative (TEI), Christopher M. Sperberg-McQueen reclamó la figura de Mercurio como patrono de la informática y abogó así, de nuevo, por la celebración del matrimonio entre Filología y Mercurio en la era del procesamiento digital de datos:

Cristopher M. Sperberg-McQueen, Pionier der *Text Encoding Initiative* (TEI), einem Konsortium, das Standards für die digitale Textdarstellung entwickelt, nimmt Merkur als den Schutzherrn der Informatik in Anspruch und feiert so noch einmal die Hochzeit der Philologie und des Merkur in der digitalen Datenverarbeitung (Sperberg-McQueen 2001: 3-4).

Desde sus más remotos orígenes, la filología se ha encargado de la elaboración de glosas, índices, glosarios, diccionarios (sema- y onomasiológicos) y otros instrumentos cuyo objetivo es la hermenéutica o interpretación de los textos. En esa característica se basa la definición de Karl Stackmann, quien entiende la filología como ciencia de los textos antiguos y de su comprensión y cuyos instrumentos son precisamente diccionarios, gramáticas, ediciones e historias de la literatura:

Eine *zweite* Option, das philologische Tun einzufangen, liegt in der Definition von Karl Stackmann [...], der in der Philologie eine "Wissenschaft" sieht, welche "Texte der Vergangenheit verfügbar macht und ihr Verständnis erschließt". [...] Wörterbücher, Grammatiken, Editionen, Literaturgeschichten tragen die Ergebnisse in die Seminare, in die Schulen, in die Öffentlichkeit. [...] Eine pure Erschließungsphilologie führt in eine transparente, aber uninteressante Welt uniformer Metadaten, eine reine Unverfügbarkeitsphilologie in eine Welt del falschen Demut, des arkanen Raunens (Lepper, 2012: 38-39).

Tras un siglo XX que ha visto el ocaso de las filologías, el comienzo del siglo XXI, de la mano de la revolución digital, ha propiciado un resurgir del interés por los métodos y técnicas filológicas¹. Proyectos de indización semántica como Wordnet, que permiten la construcción de ontologías semánticas, posibilitan la creación de índices aplicados a corpus textuales digitales. Buscadores, corpus lematizados, bases de datos léxicos y ediciones alineadas de diccionarios (como los disponibilizados por la Fundación Rafael Lapesa) están revolucionando el mundo de las aplicaciones y herramientas informáticas para trabajar con el vocabulario desde un punto de vista diacrónico y textual.

2.2. Definición de "léxico político"

Resulta muy difícil delimitar y definir qué se entiende por "léxico político" y qué términos se incluyen, o deben incluirse, dentro de él. Esta dificultad que se observa en todas las sociedades y en cualquier momento histórico se acentúa cuando se trata de comunidades muy politizadas (Hinojo Andrés, 1997-1998: 191).

Desde la concepción foucaultiana de "arqueología del saber", "la historia del pensamiento, de los conocimientos, de la filosofía, de la literatura parece multiplicar las rupturas y buscar todos los erizamientos de la discontinuidad" (Foucault, 2009 [1969]: 15), partiendo de la revisión del documento y utilizando el concepto de discurso. A partir de estas premisas, hemos tratado de superar las dificultades a la hora de definir los conceptos propios del discurso político recurriendo a la *Begriffsgeschichte* y a la antropología lingüística; pues es indudable que existe un discurso político², para cuya elaboración debe contarse con un vocabulario característico además de unos conceptos propios y exclusivos. La

¹ "Dass der Philologiebegriff zu Beginn des 21. Jahrhunderts an Attraktivität zurückgewonnen hat, liegt nicht zuletzt an Partikularentwicklungen, am Reintegrationswillen der beteiligten Fächer und Fachquerschnitte angesichts gestiegenen Konkurrenzdrucks" (Lepper, 2012: 42).

² "Unter Diskursen werden vor allem institutionell bedingte Kommunikationsbereiche und Kommunikationsformen verstanden. Einzelne Kapitel sind hier dem politischen, dem religiösen und dem Rechtsdiskurs, dem ökonomischen und dem wissenschaftlichen Diskurs gewidmet" (Kuße, 2012: 24).

antropología política nos brinda los conceptos clave de *campo* y *arena*³, espacio conceptual donde debemos situar al autor y a los traductores del *Rómulo*, que debe redefinirse como documento complejo y a la vez instrumento del discurso político dentro de la historia de las ideas y los conceptos en la que surge. En palabras de Foucault: “la historia es lo que transforma los *documentos* en *monumentos*, y que, allí donde se trataba de reconocer por su vaciado lo que había sido, despliega una masa de elementos que hay que aislar, agrupar, hacer pertinentes, disponer en relaciones, constituir en conjuntos” (Foucault, 2009 [1969]: 16-17); la elaboración de unos índices conceptuales de las traducciones del *Rómulo* sería el *vaciado* de los documentos que componen el texto u obra, en este caso, y supondrían, precisamente referida a los conceptos políticos, el aislamiento, agrupamiento, relación, jerarquización y dotación de pertinencia de esos mismos conceptos en el ámbito del discurso político como uno de los caracteres imprescindibles de la arena como campo normativizado del juego o la competencia política en el Barroco, tal como se muestra en el *Rómulo* y en sus traducciones.

3. HACIA LOS ÍNDICES CONCEPTUALES: EL *ROMULO* DE VIRGILIO MALVEZZI

3.1. Fijación de los conceptos para el etiquetado y análisis

En una primera fase de experimentación en la creación de índices conceptuales para el portal Andrés de Poza, se realizaron tentativas de adaptación parcial de la red Wordnet, en campos semánticos más o menos autónomos que permitieran por lo tanto ser escindidos del conjunto. Sin embargo, estas tentativas fueron infructuosas debido, en primer lugar, a la escasa estructuración lógica de los campos estudiados en Wordnet; y en segundo lugar, a los problemas que supone aplicar una ontología semántica contemporánea a textos clásicos con un sistema conceptual subyacente propio de su época y culturalmente condicionado.

Para el estudio del léxico político del *Romulo* se optó, por lo tanto, por establecer en primer lugar un listado de los elementos constitutivos de la arena presentes en el texto. Se podría así empezar por destacar los individuos que entran en conflicto en la obra, es decir los nombres propios mencionados⁴, que podrían ser objeto de un índice de antropónimos como los mostrados en apartados anteriores.

Una vez excluidos estos, el índice conceptual en el que estamos trabajando se ocupa de aquellas voces cuyo significado pueda estar relacionados con los elementos básicos de todo discurso político: *apoyo, élite, estado, etnicidad, facción, género, grupos de edad, institucionalización del liderazgo, jefatura, legitimidad, matriarcado, modo de producción, nacionalismo, poder, poder/saber, reglas normativas y pragmáticas, sistema político, tribu...* (Lewellen, 2009 [2003]: 301-307). Desde el punto de vista lingüístico, se trata de léxico semántico, excluidas las palabras gramaticales: fundamentalmente verbos y sustantivos abs-

³ “Un campo político es aquel en que los grupos rivales *no* comparten reglas convenidas de antemano para regular sus conflictos. Una arena política será un área en la que los competidores aceptan las mismas reglas” (Lewellen, 2009 [2003]: 127).

⁴ Amulio, Anco, Fáustulo, Fáustulo Pastor, Lucumo, Marte, Numa, Numitor, Pan, Proca, Remo, Roma, Rómulo, Tarquino, Tibre y Tulio.

tractos, con alguna presencia de sustantivos concretos y unidades pluriverbales o sintagmáticas.

Podríamos establecer, por tanto, un primer nivel conceptual de etiquetado, de acuerdo con el agrupamiento propio de la Antropología política, tal como el presentado en la tabla siguiente:

1. Apoyo	<i>adulación, aduladores, agrado, alabança, alabar, servicio</i>
2. Élite	<i>fortuna, gozar, gozo, gracia, hijos de rey, palacio, personas de sangre real, sacerdote</i>
3. Estado	<i>arte de quitar reinos, averle quitado el reino, conquista, defender, dominio, máquina del estado, ministro, privado del reino, reino, súbdito, términos reales</i>
4. Etnicidad	<i>albanos</i>
5. Facción	<i>acciones marciales, afrenta, ambición, ardimiento, arma, combatir, debilidad, deseo, despreciar, embidia, enciende, enemigo <-> amigo, engañar, engaño, escolta, espada, fiereza, flaqueza, guerra, homicidio, huir, injuriar, inquietud, miserable, miseria, ocasión, ofensa, peligro, perder, perseguidor, pusilanimidad, rendir, resistir, resistencia, recelar, seguir, temer, temeridad, temor, tumulto, vencer, venganza, violencia, violentar, violento, victoria, vituperio, voluntario, celoso</i>
6. Género	<i>cuerpo, hombre, mujer, persona, ser humano, sexo de la hembra, varón, vestal</i>
7. Grupos de edad	<i>abuelos, acciones de los antiguos, acciones de los que passaron, antiguo, años, descendencia de Remo ('ascendencia'), edad, eternidad, favor de la fama, hechos de los presentes, juventud, origen, tiempo</i>
8. Institucionalización del liderazgo	
9. Jefatura	<i>cabeças, eminencia, furor tiránico, gloria, glorioso, grandes hechos, hechos de los príncipes, imposiciones, magestad, mandar, poder que exercen los príncipes, principado, real, regir, rey, señor, sucesor, tiranía, usurpador</i>
10. Legitimidad	<i>assegurar, beneficio, bondad, desdicha, disculpa, discurso, feliz, generoso, héroe, inferior, lícito, satisfacer, seguridad, seguro, vicios del príncipe, virtud del príncipe</i>
11. Matriarcado	<i>casa del señor ('familia del señor'), donzella, hermano, hijo, mancebo, niño, padre, sobrino</i>
12. Modo de producción	<i>acontecimientos, adquisición de bienes, caso, fabricar, interés, ocio, ocupado, pastor, pobreza, riqueza, servidumbre, valor</i>
13. Nacionalismo	<i>naciones</i>
14. Poder	<i>adoración, amor del pueblo <-> aborrecimiento del tirano, atrevido, atrevimiento, autoridad, fuerza, fuerte, obedecer, obligación, obligar, poder, poderío, tolerar, voluntad</i>

15. Poder/saber	<i>alibio, aunar, bienhechor, caridad, castigo, compadecer, compasión, condescender, condenar, convenir, cruel, crueldad, dañar, daño, dichoso, dificultad, dios, educación, elemento, elocuencia, entendimiento, galardonar, guardar la razón, hado, hombres bien afortunados <-> hombres desdichados, honestar, igual por naturaleza, impiedad, impío, infelicidad, ingenio, libertad, muerte, naturaleza, necesidad, opinión, pensamiento, piadoso, piedad, prudencia, razonamiento, reputación, sagacidad, sano consejo, severa justicia de los sacerdotes, sujeto, vergüenza, vicio, vida, virtud, vocablos de los bien afortunados, yerro</i>
16. Reglas normativas y pragmáticas	<i>armas de la elección, estrupar, ética, falsario, honra, honrar, igualar, ingratitud, injusticia, inocencia, juicio, justicia, justo, ladrón, maldad, misericordia, ninguno es superior a los otros, paciencia, sacrilegio, salvar, utilidad agena, vil</i>
17. Sistema político	<i>arte con que está acostumbrado a vivir, consejo, pueblo</i>
18. Tribu	<i>aldea, ciudad</i>

Sin embargo, resulta necesario introducir aún, entre esta amplia clasificación de la antropología y el nivel del léxico (multilingüe) incluido en cada una de esas casillas, un nivel semántico en el que se situarían los conceptos⁵ –desprovistos, aún, de un ropaje léxico– del campo semántico político. En el ejemplo piloto que hemos desarrollado, aplicado a la edición múltiple de *Il Romulo* de Virgilio Malvezzi⁶, estamos interesados en las diversas acepciones del término “pueblo”, en relación con el campo político. Aunque existen más significados posibles, nos hemos centrado en tres de ellos:

- a) ‘pueblo’ como conjunto de los ciudadanos que componen un estado u otra entidad política autónoma (*demos* o *populus*);
- b) ‘pueblo’ como conjunto de individuos unidos por un origen y una cultura común (*ethnos*);
- c) ‘pueblo’ como clase social inferior opuesta a las elites dirigentes (*plebs*).

Como se puede observar, se trata de tres conceptos que, pese a poder ser encarnados –en principio– por una misma expresión léxica, se sitúan en tres casillas diferentes en la clasificación anteriormente indicada a partir de la Antropología política:

- a) ‘pueblo’ como *demos* se sitúa en el apartado de “estado”;
- b) ‘pueblo’ como *ethnos* se sitúa en el apartado de “etnicidad” o “nacionalismo”;
- c) ‘pueblo’ como *plebs* se sitúa en el apartado “sistema político”.

A partir de este nivel conceptual (*demos*, *ethnos*, *plebs*) es posible establecer en los textos la clasificación de las diversas expresiones léxicas en las diver-

⁵ En este contexto se utiliza el término “concepto” en un sentido utilitario y no histórico, como es el caso en el campo de la historia de los conceptos; se trata, por lo tanto, de unidades de contenido semántico como los empleados en ontologías o redes semánticas como Wordnet.

⁶ Las referencias completas de las ediciones empleadas para la edición digital multilingüe del *Romulo* de Virgilio Malvezzi se encuentran en la bibliografía final.

sas lenguas: fundamentalmente *popolo*, *pueblo*, *peuple* y *people*, pero también *gente*, *volgo*, *nazione*, etc. Para ello, sin embargo, es necesario desarrollar un etiquetado que responda a estos diversos niveles léxico-semánticos, y que al mismo tiempo sea respetuoso con el estándar de etiquetado TEI.

3.2. Etiquetado

Dados los diversos niveles de análisis apuntados en el apartado anterior, y a la necesidad de establecer adecuadamente las interrelaciones entre las diversas versiones multilingües de la obra, se ha desarrollado un etiquetado específico basado en la estándar TEI, creado a partir del elemento <term> con un conjunto extenso de atributos que permitan definir no sólo la forma léxica lematizada de cada ítem, sino también el concepto con el que se relaciona (*demos*, *etnos* o *plebs*), y, en el caso de las versiones traducidas, la forma léxica y el concepto subyacente en el término correspondiente del original italiano. La etiqueta en su versión completa responde por lo tanto al siguiente modelo:

```
<term xml:id="t0001" xml:lang="es" key="pueblo" n="c001" sortKey="popolo"
      type="c001">
```

- @xml:id es un valor único en cada uno de los textos, que permite identificar cada *token* etiquetado y establecer la relación entre lenguas;
- @xml:lang indica mediante un código estándar la lengua del documento etiquetado;
- @key establece el lema de la palabra etiquetada;
- @n indica el concepto con el que se vincula esa aparición de la palabra (siendo "c001" *pueblo* = *demos*; "c002" *pueblo* = *plebs*, y "c003" *pueblo* = *etnos*);
- @sortKey establece la relación del término traducido con el lema del original italiano (y aparece, por lo tanto, en todas las versiones excepto el original);
- @type indica el concepto con el que se relaciona el término en el original italiano (lo que permite percibir con mayor claridad si se ha producido una modificación semántica durante el proceso de traducción).

3.3. Soluciones de visualización

Para la visualización del índice conceptual, y siempre teniendo presentes los potenciales intereses de los usuarios-lectores, se ha optado por una solución que permita percibir la jerarquización de niveles anteriormente mencionada⁷. Así, el índice conceptual desarrollado se estructura en tres columnas:

- En la primera columna, se establecen los tres conceptos diferentes anteriormente definidos: *pueblo* = *demos*, *pueblo* = *plebs*, *pueblo* = *etnos*.
- En la segunda columna se ofrecen las manifestaciones léxicas de cada uno de los tres conceptos, en el original italiano.
- En la tercera columna se ofrecen las traducciones de esas manifestaciones léxicas del original, en las restantes lenguas, independientemente de que

⁷ El desarrollo informático necesario para la visualización del índice ha sido realizado por la empresa Bitext.

mantengan el sentido del original o lo sustituyan por otro (más adelante volveremos sobre este aspecto).

Conjunto de gentes (19)	gente (4)	en: people (3)
		es: gente (4)
		fr: people (2)
	popolo (15)	en: people (14)
		en: common people (2)
		es: pueblo (15)
		es: hombre (1)
		fr: people (12)
		fr: homme (1)
		fr: nation (1)

Fig. 3. Índice conceptual para el concepto 'demos'.

El usuario-lector puede, a continuación, ver todas las apariciones de determinada expresión léxica en una de las lenguas (pinchando encima de cualquiera de las expresiones que aparecen en la columna segunda o tercera), o bien, si lo prefiere, obtener una visión de conjunto de todos los casos relativos a uno de los conceptos en todas sus expresiones léxicas multilingües, clicando para ello en el concepto que aparece en la primera columna.

Venne sospitione al popolo , che i Senatori (quali haueua spogliati dell'autorita) l'haussero amazzato.	Le peuple soupçonne que les Sénateurs dégradés par Romulus de l'autorité qu'il leur auoit luy mesme donnée luy auoient fait oster la vie.	Sospeché el pueblo que los senadores, a quien avia quitado la autoridad, le avian muerto.	The people did suspect that the Senators (whom he had deprived of their authority) had slaine him.
Tumultua il popolo . Alza il bollore, ma non lo spande fuori del vaso: Si mostra pronto a seguire chi voglia vendicare. Un Senatore, che in quel punto si fosse fatto capo del popolo , si facoua, non hà dubbio, Signore della Città.	La mort de Romulus esmeut le peuple , il boult de tumulte, mais il ne sort pas pour tant pardessus les bords du vaisseau: Il ne demande qu'à remuer les mains, et disent qu'ils sont tous prêts à sulure celuy qui voudra la venger. Le beau jeu qu'il y auoit alors pour vn ambitieux! Un Sénateur qui se fust à ceste là déclaré chef du peuple , se fust sans doute fait chef et maistre de la Cité.	Alberótase el pueblo , hierbe, mas no vierle fuera del vaso el herbor: muéstrase pronto a seguir al que quisiere vengança. Un senador que en aquella ocasión se huviera hecho cabeça del pueblo , se huviera hecho, sin duda, cabeça de la ciudad.	The people make a tumult; their choler boyles apace, but leapes not out of the vessel; they shew themselves ready to follow any who will revenge him; some Senator, who at that time had mado himselfe the peoples Leader, might doubtlessely have made himselfe Master of the Cité.
Giulio Proculo vi accorre, raccontando hauer veduto Romulo ascendere al Cielo, e che comandaua di voler essere chiamato Dio Quirino. Il popolo lo crede; l'acqueta, ed in vece di vendicarlo, si dà a sacrificargli.	Julius Proculus suruint là dessus qui appaisa toute la sedition, jurant et protestant qu'il auoit veu Romulus enleuer dedans le Ciel, et qu'il luy auoit commandé de dire à son peuple qu'il honorast de sormalis sa memoire du nom de Quirin. Le peuple adouesta foy au rapport de Proculus, parce qu'il estoit estimé fort homme de bien: De façon qu'au lieu de rechercher la vengeance de la mort de Romulus, ils luy voient des sacrifices comme à vn Dieu.	Julio Préculo los securrió afirmando que avia visto subir al cielo a Rómulo, y que mandava que le llamasen Dios Quirino. El pueblo lo cree, se quieta y en lagar de vengarle le sacrifica.	Julius Proculus came in, affirming that he saw Romulus ascend up into heaven, and that his will was to be called the god Quirinus; the people believe him, appease the tumult, and in stead of revenging him, prepare to offer sacrifice unto him.

Fig. 4. Expresiones léxicas multilingües del concepto 'demos'.

Mediante esta comparación de las diversas traducciones de un determinado concepto en el original italiano, es posible analizar con especial claridad las elecciones verbales de los traductores, o las diferencias entre las diversas lenguas en su estructuración del campo semántico y de sus representaciones léxicas. Así, por ejemplo, se observa que Henry Carey, traductor de la versión inglesa, enfatiza sistemáticamente el concepto *pueblo* = *plebs* mediante la lexía compleja *common people*.

<p>Gli stati, che dormono quieti, perche sono amici de' vicini, hanno gran ventura, se incontrano in qualche occasione di sdegno; e g'huomini auertiti, in cosi fatti casi la cercano, perche il popolo non si lascia persuadere se non quello che vede. Egli giudica col'occhio, non coll'intelletto; ne vi hà argomento presso di lui bastevole, per espugnare l'apparenza. Il tenere amicia de' vicini è ottimo. Sopra di quella fondare la sicurezza dello stato è pessimo. Stanno bene, per amici, se si considerano per nemici, acciò debbano amare, e non possono offendere. L'altezza di quell'edificio, che piace, quando altri crede, che g'habbia a seruire d'habitatione, dispiace, se vi si aspetta il precipitio.</p>	<p>Les Estats qui dorment en repos, à cause qu'ils sont amis de leurs voisins sont fort heureux quand ils rencontrent quelque quiete de courroux: Les hommes bien auidsez le cherchent tousiours en ces temps là; parce que le peuple ne peut jamais apprehender que ce qu'il voit present deuant soy; il iuge avec la veuë et non pas avec l'entendement; Il n'y a point d'argument assez fort qui luy puisse contester l'apparence. Il est bon d'auoir amitié et alliance avec l'Estre voisin, mais c'est faire mal de fonder la seureté du sien là dessus: nos voisins sont bons pour amis, quand nous les considerons comme s'ils estoient ennemis, afin qu'ils aiment seulement, et qu'ils ne puissent offenser: La grandeur et hauteur d'un edifice est tres-agreable à celui qui croit d'en faire son habitation, mais quand il preuoit que cette hauteur là doit estre son precipice, elle luy est odieuse.</p>	<p>Los estados que duermen quietos, porque son amigos de los vezinos, bienen gran dicha si encuentran en alguna ocasión de enojo; y los hombres advertidos en semejantes casos la buscan, porque el pueblo no se dexa de persuadir sino lo que ve. El juzga con la vista, no con el entendimiento; ni ay argumento eficaz para él que le contraste la apariencia. El tener amistad con los vezinos es bueno. Sobre aquella fundar la seguridad del estado es malo. Son buenos para amigos si se consideran por enemigos, para que deban amar, y no puedan ofender. La altura de aquel edificio que agrada quando uno cree que le ha de servir de habitación, le aborrece quando le considera como precipicio.</p>	<p>States that enjoy rest and quiet, because they are in league and friendship with all their Neighbours, have great good fortune if they meet with any occasion of offence; and wise men in such like occasions seeke after them; for the common people will not be perswaded to more than what they see, they judge by the eye, not the understanding, nor is there any argument with them of force enough to gain-say appearance. To keep friendship with neighbours is exceeding good; but hereupon to build the securitie of a State, is passing bad: they are well held for friends, if they bee considered as enemies; that so they may bee bound to love, and not able to hurt; the height of that building, which is delightfull and pleasing, when one thinks to enjoy it for an habitation; is displeasing, if he looke it shall fall upon his head.</p>
--	--	--	--

Fig. 5. Ejemplo de selección léxica para 'pueblo = plebs'.

Por otro lado, resulta también de vital importancia indicar, en la visualización que se ofrece al lector, aquellos casos en los que se ha producido una alteración de conceptos en el proceso de traducción, mediante un código de colores: el azul indica que existe identidad entre los conceptos en las diversas lenguas; el color verde indica que se ha producido una sustitución de un concepto por otro. En el ejemplo que se muestra a continuación, el original italiano "il popolo", que hace referencia muy probablemente al conjunto de los ciudadanos (*demos*), es sustituido por "common people" (*plebs*) en la traducción inglesa.

<p>Hà il Principe vn non sò che più dell'huomo, nella maestà del volto, ne g'Angli, che lo difendono, nelle stelle, che g'influiscono. Alcuni gli diedero nome d'Heroe, la verità la trauati dai retto, se equiuocando dalla simiglianza all'essenza, al nome di Dio non haussero aggiunta l'adoratione. Il popolo, perche lo crede maggiore dell'huomo, si marauiglia se lo vede uguale, si scandalizza se lo conosce inferiore. Non deeno i Prindpi lasciarsi misurare. Cimentarsi al paragone senza sicurezza di vincere, è sicurezza di perdere. Vn non sò che di più, che da g'altri si desidera, do chi hà un non sò che di più de g'altri.</p>	<p>Les Princes ont tousiours quelques graces particulieres par dessus les autres hommes, soit en la maesté de l'aspect, soit que les Anges les defendent, ou bien en ce que les estoiles versent sur eux leurs benignes influences. Aucuns les ont nommez Heros, autres diuins, et les Gentils n'eussent pas excocé ce qui est licite, si prenant la ressemblance pour l'essence, ils n'y eussent adouste l'adoration. Les hommes s'esmerueillent de les voir imperieux comme les Dieux, et se scandalisent quand ils les voyent inferieurs: Les Princes ne se doiuent point laisser mesurer: se rualer à la comparaison sans assurance de vaincre, on se doit assureur de se perdre: Il est necessaire qu'il y ait vn le ne scay quoy plus qu'aux autres, en celuy qui a vn le ne scay quoy par dessus les autres.</p>	<p>Tiene el príncipe un no sé qué más que hombre en la magestad del semblante, en los ángeles que le defendien, en las estrellas que le influyen. Algunos le dieron nombre de héroe, la verdad lo llamó dios; y los gentiles no hubieran excodido de lo licito si, equiuocando la semejanza a la esencia, al nombre de dios no hubieran justado la adoración. El hombre, porque lo ve mayor que el hombre, se marauilla si le ve igual, se escandaliza si le conoce inferior. No deben los príncipes dexarse medir. Baxarse a la comparación sin seguridad de vencer, es seguridad de perderse. Un no sé qué más que en los otros se desea en quien tiene un no sé qué más que los otros.</p>	<p>Princes have somewhat more than other men in the majestie of their countenance, in their tutelary Angels, and in the influences of the Starres. Some call a Prince an Heroe, Truth it selfe calls him a god, nor had the Gentiles benee much out, had they not by equiuocating from the similitude to the essence, added to the Name of God adoration; the common people, because they thinke him more than man, wonder if he be but equal to other men, are offended if he be inferior: Princes ought not suffer themselves bee measured; who suffers competition, not sure to win, is sure to lose; somewhat of more than from others is required from him that hath somewhat of more than others.</p>
--	--	--	--

Fig. 6. Sustitución de conceptos en la traducción.

Naturalmente, serían necesarias, para extraer conclusiones definitivas respecto a la configuración de los campos semánticos implicados, investigaciones más exhaustivas que tuviesen en cuenta los condicionantes lingüísticos, culturales y políticos de cada traducción, empleando para ello, entre otras fuentes documentales imprescindibles, los diccionarios históricos correspondientes a cada una de las lenguas.

4. CONCLUSIONES: POSIBLES UTILIZACIONES FUTURAS DE ESTOS ÍNDICES

Lo que hemos mostrado en las páginas anteriores es sólo una prueba piloto de creación de un índice conceptual, muy limitada por lo tanto en su alcance y pretensiones, con un reducido núcleo de conceptos interrelacionados pertenecientes al ámbito político. Con todo, y a pesar de lo limitado de la muestra escogida, la experimentación en el establecimiento de un conjunto de conceptos, su etiquetado y visualización han permitido extraer conclusiones que serán de gran utilidad para futuras experiencias en esta misma línea.

Así, la ampliación de esta experiencia piloto a un corpus más amplio, y al conjunto del léxico político europeo del siglo XVII, exigiría la creación de una red de conceptos o de una ontología propia, dado que, como hemos visto, las ontologías genéricas actualmente existentes (como WordNet) tienen obvias carencias cuando se aplican a campos semánticos que quizá no estén todavía suficientemente estandarizados, y en especial cuando se aplican a textos históricos y no contemporáneos. Es este un proceso que, como ya se ha apuntado, exigiría un riguroso trabajo filológico, que combinase una aproximación desde la antropología política, con una investigación lexicográfica multilingüe y contrastiva.

En cualquier caso, un desarrollo como el apuntado aquí, en particular si se extendiese a todo el corpus Andrés de Poza, o incluso a un corpus más amplio (sincrónico o diacrónico) permitiría realizar estudios en áreas diversas, tales como los estudios de traducción; los estudios de lexicografía; la historia de los conceptos o la antropología política. La aplicación de técnicas informáticas de tratamiento y enriquecimiento de textos se alía así, una vez más, con los objetivos de la filología en su sentido más amplio y más fructífero.

BIBLIOGRAFÍA

- Foru Ondarea, <<http://foruondarea.net/>> [10/11/2013].
 Foucault, Michel, *La arqueología del saber*, Madrid, Siglo XXI, 2009 [1969].
 Fundación Rafael Lapesa, <<http://www.frl.es/Paginas/default.aspx>> [10/11/2013].
 Hinojo Andrés, Gregorio, "El léxico político romano: *speciosa uerba*", *Voces*, 8-9 (1997-1998), pp. 191-206.
 Kuße, Holger, *Kulturwissenschaftliche Linguistik. Eine Einführung*, Göttingen, Vandenhoeck & Ruprecht, 2012.
 Lepper, Marcel, *Philologie zur Einführung*, Hamburg, Junius Verlag, 2012.
 Lewellen, Ted C., *Introducción a la Antropología Política*, Barcelona, Edicions Bellaterra, 2009 [2003].
 Malvezzi, Virgilio, *Il Romulo*, Bologna, Clemente Ferroni, 1629.

Malvezzi, Virgilio, *Romulus and Tarquin*, traducción de Henry Carey, Londres, J. Haviland, 1637.

Malvezzi, Virgilio, *Le Romulus du marquis de Malvezzi, avec ses considérations politiques et morales sur sa vie*, traductor anónimo, Paris, Paris J. LeBouc, 1645.

Malvezzi, Virgilio, *El Rómulo*, traducción de Francisco de Quevedo, edición crítica de Carmen Isasi, Bilbao, Universidad de Deusto, 1993.

MultiWordNet, <<http://multiwordnet.fbk.eu/online/multiwordnet.php>> [10/11/2013].

Portal Andrés de Poza, <<http://www.andresdepoza.com>>.

Sperberg-McQueen, Christopher M., "Die Hochzeit der Philologie und des Merkur: philologische Datenverarbeitung", en Stephan Moser et al. (eds.), *Maschinelle Verarbeitung altdeutscher Texte. Beiträge zum 5. Internationalen Symposium 1997*, Tübingen, Niemeyer, 2001, pp. 3-22.

Text Encoding Initiative, <<http://www.tei-c.org>> [10/11/2013].

Wordnet Multilingual Central Repository Interface, <<http://adimen.si.ehu.es/cgi-bin/wei/public/wei.consult.perl>> [10/11/2013].



RESUMEN

En los progresos de la edición digital de textos literarios, la elaboración de índices, tesauros y glosarios ofrece todavía un amplio espacio de exploración. En el marco del proyecto Andrés de Poza (<<http://andresdepoza.com>>), dedicado a la edición de textos literarios europeos de los siglos XVI y XVII en versiones múltiples (original y varias traducciones) se ha trabajado ya en la preparación de índices de diverso tipo. En el momento actual, se ha considerado la posibilidad de integrar índices temáticos, sustentados en los propios índices de materias de los testimonios editados, o bien de nueva creación. En cuanto a la resolución técnica, contamos en este sentido con la experiencia obtenida en los índices temáticos de las *Cartas Puebla de Bizkaia* (<<http://foruondarea.net>>), elaborados a partir de una marcación XML.

Este ensayo presenta las decisiones tomadas a lo largo del proceso de creación del índice conceptual multilingüe a través de un ensayo realizado con vocabulario del campo político (un ensayo limitado pero metodológicamente exportable a otros términos del mismo campo, así como a otros campos). Finalmente, se intenta también mostrar con una breve ejemplificación las potenciales aplicaciones futuras de este tipo de índices: estudios de léxico histórico, estudios de traducción, estudios culturales comparados y sobre todo historia de los conceptos –en la línea de la antropología política–, además de su propio interés como herramienta de acceso a los textos multilingües editados.

Palabras clave: Índices, marcación, edición, textos alineados, léxico político.

ABSTRACT

In the last developments of literary texts digital editing, the making of indexes, thesauri and glossaries still offers an open field for researching. In the frame of the project "Andres de Poza" (<<http://andresdepoza.com>>), devoted to European sixteenth and seventeenth centuries literary texts editions in multiple versions (original and translations to several languages), we have already worked on the preparation of indexes of various kinds. At the moment, we have considered the possibility of integrating thematic indexes for the edited or newly created textual witnesses, supported by their own tables of contents. As for the technical resolution, we count on the experience obtained in the subject indexes of the *Cartas Puebla de Bizkaia* (<<http://foruondarea.net>>), elaborated with an XML marking code.

This paper presents the decisions made throughout the process of creating multilingual indexes of the vocabulary from the political lexicon (one methodologically limited trial but exportable to other terms in the same semantic field, as well as to other fields or lexicons). Finally, we show also potential future applications of this type of indexes: diachronic studies of the vocabulary, translation studies and traductology, comparative cultural studies and especially history of concepts –as meant in political anthropology–, as well as a tool for accessing multilingual edited texts and corpuses.

Keywords: Indexes, marking, editing, text alignment, political lexicon.

The Littera Project: *Webpapers* o trabajos académicos 2.0

MÓNICA POZA DIÉGUEZ
Hobart and William Smith Colleges
pozadiequez@hws.edu

En el año 2010 Sidonie Smith, por entonces presidenta de la Modern Language Association, reflexionaba sobre las transformaciones que se están dando en el plano académico como consecuencia de la implantación de los medios digitales. Smith abordaba concretamente el espinoso asunto sobre el formato del trabajo académico e investigador, en un momento en el que la adquisición, la comunicación y la evaluación del conocimiento se ligan por completo a nuevos canales y medios:

“Digital media and computational technologies are radically transforming how knowledge is produced, communicated, and evaluated. The digitalization of scholarly work in the humanities brings new modes of research; new formats of presentation; new networks for communication; and new platforms for organizing knowledge, orchestrating argument, and visualizing intellectual exchange. Doctoral students in the modern languages will increasingly create and use digital archives and invent multimodal forms of scholarly presentation and communication in the next decade. Why should the dissertation remain inflexibly wedded to traditional book-culture formats?” (Smith, 2010: 2-3).

De alguna manera, la presidenta de la MLA ponía sobre el tapete una cuestión sobre la que se llevaba debatiendo ya hace varios años en numerosos medios, en torno al impacto que la tecnología digital e internet producen en el ámbito de la educación. A los debates acerca de si las revistas científicas y de investigación deben continuar con su publicación impresa en papel, o si deben distribuirse en plataformas o a través de redes sociales (al tiempo que se testan nuevos modelos de suscripción), se suman aquellos otros que tienen como centro de interés la dinámica de la enseñanza en el aula. Así, pululan por la red innumerables artículos de todo tipo en los que se hace evidente la inquietud del cuerpo docente ante el desafío que implica la era digital. El problema reside, aunque sin generalizar, en el hecho de que un nutrido grupo de profesores esté todavía más acostumbrado a usar internet como herramienta diaria en sus quehaceres cotidianos, que a incluir en sus clases formatos o medios no tradicionales, tales como ordenadores portátiles, tabletas, teléfonos móviles, blogs, redes sociales, etc., como si el modelo de la clase tradicional no pudiera incorporar las mismas herramientas que se usan fuera del aula y de manera cotidiana.

Y sin embargo, progresivamente sucede que aumenta el porcentaje de alumnos que desde la escuela se familiarizan con una educación mediada por las pizarras digitales, los laboratorios de informática, el uso de internet. Por tanto, el desafío consiste en conectar al profesorado con una generación acostumbrada a una comunicación abierta, colaborativa, multimodal y transmedia. En este sentido, parece que los docentes de educación primaria y secundaria son menos reticentes a la incorporación de un lenguaje más comunicativo que posibilite el intercambio de conocimiento entre quien enseña y quien aprende. Existen así proyectos educativos que demuestran este nuevo entorno de trabajo. Uno de ellos es, por ejemplo, el proyecto llevado a cabo por Sergio Tejero, profesor de educación secundaria en el Instituto Ekialde, en Vitoria, ampliamente difundido en varios medios de comunicación como el Telediario de Televisión Española¹. Los objetivos del proyecto consistían en que los estudiantes aprendiesen y practicasen el resumen, al tiempo que asimilaran el uso responsable de las redes sociales. Para ello, se tomó una obra clásica como *El ingenioso hidalgo don Quijote de la Mancha*, que los estudiantes resumieron capítulo a capítulo, en 140 caracteres, en "tweets", a través de la cuenta @ElQuijoTweet, en la conocida red social Twitter.

Asimismo, existen presentaciones en Slideshare desarrolladas tanto por estudiantes como por profesores sobre los más diversos temas. También la red social de videos caseros YouTube ha abierto un espacio diferente y nuevo en el que los estudiantes exploran las diversas materias, en calidad de receptores (y aquí habría que mencionar la Khan Academy), pero también en el que ejercitan su rol como autores al aportar nuevo contenido a la red. La lista de programas y herramientas de interacción social que pueden usarse en la educación es ingente, y cada día surge una multiplicidad casi imposible de rastrear.

Es obvio que nos encontramos ante un cambio de paradigma que proviene del desarrollo tecnológico, y que afecta al soporte en el que hasta ahora hemos depositado el conocimiento. Si ya el cambio de una cultura oral a una cultura de lo impreso constituyó un paso decisivo en la evolución y el desarrollo de la civilización occidental, el momento en que vivimos no supone sino una vuelta de tuerca más en ese desligar el pensamiento humano de nuestro propio cuerpo para depositarlo en un objeto externo al cerebro, antes en un libro y ahora en un "disco duro externo". Por tanto, es obvio que existe un vínculo entre ese soporte físico artificial, esa memoria exterior, y la adquisición de conocimiento. Pero si bien la cultura analógica, basada en el soporte del libro, tiende al cultivo del aprendizaje individual con una experiencia lectora sólo posible de compartir *a posteriori*, el uso de internet y los diferentes soportes que la industria tecnológica ofrecen hoy en día contribuyen a crear una experiencia completamente diferente. Tal y como menciona Dolors Reig:

"El estilo cognitivo de los jóvenes es diferente y resulta cada vez más evidente que leemos de forma distinta, con intervalos de atención reducidos. Cada día está más obsoleto el concepto de lectura lineal y aumentan en importancia poder explorar con mayor profundidad o alejarse para volver cuando se quiera a cada concepto, posibilidades, entre otras, del hipertexto. Predomina la cultura de 'párrafos cortos' y las posibilidades de interacción, de participación, son especialmente valoradas" (Reig, 2013: 32-33).

¹ <<http://www.rtve.es/alacarta/videos/telediario/td-quiote/1661356/>>.

Es así que cobran sentido términos como el conocimiento abierto y, en consecuencia, la educación abierta basada en un modelo en el que el compartir se vuelve clave esencial del proceso de aprendizaje y formación, tal y como explica Joseba Abaitua:

“La forma de construir y compartir el conocimiento se transforma a pasos agigantados y muchos de los modelos heredados del pasado están quedando en desuso. Es una revolución inevitable y fenómenos como Wikipedia, Public Library of Science (PLoS), Comprehensive Knowledge Archive Network (CKAN), Massive open online course (MOOC) no tienen marcha atrás” (Abaitua, 2013).

Esta transición a lo digital en la que vivimos pone de relieve varios aspectos. Uno de ellos, el más importante a mi entender, es la a menudo tan mencionada desconexión entre pedagogía y medios tecnológicos, que surge en gran parte por el hecho inevitable de la brecha generacional existente entre aquellos docentes que han sentido forzada su entrada en el mundo digitalizado, y aquellos que por razón de edad se han adentrado en esta tendencia de manera más natural. La desconexión entre pedagogía y tecnología produce, en consecuencia, el aberrante efecto del aula “frankenstein”, al dotar a las instituciones académicas de todo tipo de materiales electrónicos forzando así al profesor a un reciclaje tecnológico, aunque no necesariamente pedagógico. Así, el docente suele inclinarse a usar el aparataje en sus clases tradicionales y de forma tradicional, lo que constituye un error epistemológico. Profesor y alumnos suelen rechazar este tipo de experiencias que en algunos casos pueden dar al traste con el mejor de los cursos que, de haber sido impartido de manera tradicional hubiera logrado sus objetivos sin provocar ningún trauma. Además, el curso tradicional que abusa de lo tecnológico, ese aula “frankenstein”, no supone avance alguno hacia una educación de índole digital.

Es decisivo, por tanto, debatir, proponer modelos pedagógicos que sintonicen con aquellas cualidades que le son más caras al nuevo entorno tecnológico. Para empezar, deberíamos plantearnos si la educación debe continuar siendo un pilar social parcialmente desconectado de otros ámbitos, si la universidad debe seguir siendo aquella fortaleza medieval que producía, atesoraba y custodiaba el saber, al tiempo que lo circunscribía a unas pocas élites o si, por el contrario, pueden tenderse lazos que amplifiquen la experiencia educativa y la globalicen adecuadamente pues, no en vano, durante siglos la metodología empleada por los gremios se basó en el “aprender haciendo” y, no en vano, el modelo científico se asienta sobre la pauta de la experimentación, en la que prueba y error constituyen el modelo de aprendizaje. Actualmente, resulta inconcebible imaginar un sistema en el que el modelo económico y el educativo no estén estrechamente enlazados pues, repito una vez más, estamos ante un cambio de paradigma en el que todo se conecta. Como ejemplo quisiera destacar la consultora THP², liderada por María Jesús Salido, cuyas propuestas a nivel empresarial conectan perfectamente con algunas de las propuestas que desde The Littera Project hacemos para la educación.

Algunas de las características que Salido propone para la detección de lo que se puede considerar como una empresa madura, que funciona conforme a las características de una comunidad de práctica, coinciden punto por punto con aquellas otras

² Puede consultarse el sitio web en la siguiente URL: <<http://www.theproject.ws/>>.

que son beneficiosas a la hora de trabajar en el aula 2.0. Así, frente a la empresa que premia la obediencia, está la que premia la iniciativa, frente a la imposición de uniformidad surge la empresa del siglo XXI en la que singularidad forma parte de las cualidades más destacadas. La estructura jerarquizada se sustituye por una estructura horizontal, y frente a premiar a quien sabe, en la empresa moderna se premia al que comparte. También en el mundo empresarial tecnológico, el liderazgo se basa en el servicio y no en el poder, se premia la acción y no se castiga el error³, y todo se publica porque "todo pertenece a todos" (Salido, 2012: 43). De hecho, la propia Salido y el equipo de THP trabajan así en su consultora.

En el caso concreto de las humanidades creo que se deberían proponer modelos ambiciosos, en el que los estudios humanísticos salgan de las aulas y se vinculen a las instituciones, empresas, a los ciudadanos, para lograr un impacto social y contribuir *de facto* al desarrollo y a la innovación. Y para ello es preciso adentrarse en nuevas dinámicas pedagógicas. Por tanto, lo que aquí defiende es un modelo educativo basado en el "aprender haciendo", tal y como defiende Ken Robinson⁴, una experiencia pedagógica abierta y colaborativa, que conecta individuos y organizaciones, que borra las fronteras y se propone multidisciplinar; una educación en la que el docente asume un papel de "gestor", "orientador" que no sólo asiste pedagógica y técnicamente, sino que además indica posibilidades, errores, organiza contenidos y sinergias, más que desempeñar un rol autoritario como garante de verdades incuestionables. En ese nuevo aula, el estudiante no puede formar parte de una estructura jerarquizada verticalmente, sino que debe pertenecer a un equipo integrado por él mismo, sus compañeros y el profesor, como nodos integrantes de una estructura horizontal, que, a su vez, tienen otras conexiones dentro y fuera del aula creando así una red de conocimiento. En una estructura de este tipo, la clave para desarrollar una pedagogía digital consiste en "conectar" e "intercambiar", por lo que la dinámica en la clase nunca puede ni debe ser pasiva, con estudiantes que únicamente desarrollan un rol de receptores, sino activa, en la que el curso se va tejiendo paso a paso con la colaboración de todos los integrantes.

Este modelo no es exactamente nuevo ni exclusivo al ámbito académico. Lo que sí es novedoso es su implementación extraordinaria por medio de la tecnología, que no sólo permite la deslocalización de los integrantes del grupo, sino que posibilita un proceso de trabajo muchísimo más enriquecedor, al tenerse acceso a una ingente cantidad de datos, de materiales como nunca antes se había conocido. Que dos estudiantes situados en continentes diferentes del planeta puedan "chatear" acerca de un manuscrito del siglo XV digitalizado por la Biblioteca Nacional de España, y que puedan debatir asuntos de la trama narrada en conexión con música o films alojados en internet, o que puedan analizar ese texto, y escribir un ensayo a 4 manos en Google Drive, apoyándose en artículos de investigación accesibles para ambos a través de JSTOR, nos da una idea de las posibilidades que encierra una educación expandida hasta estos límites.

³ Véase lo que Shannon Christine Mattern acerca de los "errores" como fuente de aprendizaje: <<http://www.wordsinspace.net/wordpress/2012/08/28/evaluating-multimodal-work-revisited/>>.

⁴ Dejo aquí los enlaces a las famosas charlas de Sir Ken Robinson en el TED. La mítica conferencia de 2006 puede verse aquí: <http://www.ted.com/talks/lang/es/ken_robinson_says_schools_kill_creativity.html>. Y la más reciente de 2013: <http://www.ted.com/talks/lang/es/ken_robinson_how_to_escape_education_s_death_valley.html>.

Tras esta contextualización y regresando a las ya citadas reflexiones de Sidonie Smith a propósito del formato de la tesis doctoral (entendiendo que el formato del trabajo académico no es sino el resultado, la expresión, de un modelo pedagógico concreto), le traslado al lector las siguientes cuestiones. ¿Es posible extender o expandir el formato del trabajo académico para adecuarlo a la era digital? En definitiva, ¿podemos amplificar adecuadamente las humanidades en el marco tecnológico de la era digital? En las siguientes líneas me propongo presentar The Littera Project, un proyecto docente que inicié en 2010 junto con el profesor Álvaro Llosa Sanz, como un ejemplo más de lo que puede ser un aula humanista 2.0.

The Littera Project⁵ surgió como experiencia piloto en 2010, gracias a una colaboración llevada a cabo con el Departamento de Hispanística, en la Universidad de Szeged (Hungría). A lo largo del curso académico de 2010-2011 se realizaron 4 cursos a nivel de Máster. Posteriormente, se comenzó a trabajar con universidades americanas en Estados Unidos. En el curso 2011-2012 se impartieron cursos de nivel subgraduado en Syracuse University (uno de ellos se repitió en 2013). Y en 2012 se incorporó a la lista de instituciones Hobart and William Smith Colleges con un curso más en el año académico 2013-2014.

La lista de colaboradores también se ha ido ampliando. Profesores de Syracuse University (Alejandro García Reidy), de la Universidad Complutense de Madrid (Ángel García Galiano) han realizado varias actividades puntuales como hangouts o videoconferencias por skype, o proyectos más consolidados, como sucede en el caso de Álvaro Llosa Sanz y Fernando Rodríguez Mansilla, que imparten dos secciones del mismo curso, asociados a The Littera, desde Hobart and William Smith Colleges. Asimismo la lista de colaboradores ajenos al mundo académico se ha expandido al mundo de la empresa gracias a la colaboración de María Jesús Salido y el equipo de THP, pero también al de los viajes, gracias a la conexión con el "travel blog" El Guisante Verde⁶, reseñado por Paco Nadal de El País como uno de los 25 mejores blogs de España sobre viajes. Por The Littera Project también han pasado colaboradores fugaces, como el escritor Juan Gómez Bárcena o Andrés Neuman.

Las líneas básicas de The Littera Project consisten en el desarrollo de una dinámica colaborativa en la clase, la multimodalidad, la transmedia, la realización de proyectos con impacto social, el alfabetismo digital, la comprensión y el uso de los materiales registrados bajo licencia de Creative Commons y el aula abierta. La dinámica colaborativa se lleva a cabo mediante la realización de actividades y proyectos. Las actividades deben diseñarse conforme a los objetivos concretos del curso, aunque deben estar también pensadas para facilitar a los estudiantes una progresiva familiaridad con la tecnología, pues en la mayoría de los casos no sólo desconocen una gran parte de los programas, sino que ni siquiera intuyen la forma en como se pueden usar todas las herramientas y plataformas en el plano académico.

De manera paulatina y natural, los alumnos adquirirán también, gracias al desarrollo de estas actividades, el hábito de asumir diferentes roles en la clase, con lo que se acostumbrarán a liderar grupos de investigación, editarán trabajos académicos o presentaciones, buscarán materiales adecuados en internet y en

⁵ La página web de nuestro proyecto docente es: <<http://www.thelitteraproject.weebly.com>>.

⁶ <<http://www.guisanteverdeproject.com/>>.

la biblioteca, y generarán ideas críticas o de resolución de problemas o errores. Obviamente, la evaluación en un aula de estas características no puede realizarse mediante exámenes estandarizados, sino que las calificaciones deben responder a la peculiaridad de la dinámica de grupos, de comunidad, y debe ser continua. El aula colaborativa conecta al estudiante con sus compañeros, pero también lo conecta con su profesor y con agentes externos al curso, por lo que será fundamental inculcar el intercambio de ideas, de opiniones, de información y materiales, todo ello susceptible de ser evaluado como algo más que una mera "participación" en la clase, ya que constituye la esencia misma que hace funcionar el curso.

Algunas de las actividades que se han realizado en The Littera Project han tenido como protagonistas esenciales a los blogs. A veces como espacio en el que analizar un texto y en el que familiarizarse con la práctica de dar *feedback* por medio de comentarios, pero también como CMS (Course Management System)⁷. También han sido fundamentales las presentaciones colaborativas *online* tipo PPT (Power Point)⁸, lo que nos llevó a Google Docs y su paquete de servicios integrados para elaborar en línea todo tipo de documentos, formularios, hojas de cálculo y presentaciones.

De entre todas las actividades destacaría una, el examen *online* de Mid-Term, de tipo "to take home" (hecho en casa) realizado por los estudiantes del curso de iniciación a las literaturas hispánicas, *Literatura en acción*, en Syracuse University en el año 2013, en Google Docs. El examen contenía dos partes. En la primera, el estudiante debía elaborar un ensayo analítico sobre una de las obras que se habían estudiado en clase. Este ensayo debía redactarse conforme a las normas de la MLA mostrando una bibliografía citada en su elaboración. La segunda parte del examen consistía en tres preguntas, dos de las cuales tenían como referencia un hangout⁹ que se había desarrollado durante el curso, que versaba sobre el episodio de los molinos en la obra cervantina *Don Quijote de la Mancha*, y en el que habían participado profesores y estudiantes de Syracuse University y Hobart and William Smith Colleges. Este hangout o videoconferencia múltiple se había retransmitido previamente en directo a través de la Comunidad que la clase tenía en la red social de Google conocida como Google+. Allí, los estudiantes asistentes dejaron comentarios, preguntas a las que los participantes procuraban responder en tiempo real. Posteriormente, la videoconferencia quedó alojada en YouTube, donde más público la visualizó. Para el examen, el hangout consistió en una prueba de comprensión oral para los estudiantes, que debían visualizar o revisualizar la conferencia para responder a las preguntas planteadas.

La realización del examen de medio semestre en Google Drive facilitó una interacción muy estrecha entre estudiante y profesor por medio de comentarios, notas al pie y chats. A los estudiantes se les proporcionaba una hoja de clave de errores¹⁰ para comprender los símbolos que le indicarían qué tipo de error

⁷ Véanse los cursos impartidos en la Universidad de Szeged (Hungria): <<http://fantasia-y-poder.weebly.com/>>, <<http://cuentohispano.weebly.com/>> y <<http://masalladelpapel.weebly.com/>>.

⁸ Ejemplo de presentación colaborativa: <<https://docs.google.com/presentation/d/1CQcEEKHgN5OeEhrH5tTtLZK-Sj912ffbudGaUX2Y7-o/edit?usp=sharing>>. Véanse las figuras 1 y 2 para ver las presentaciones en su interior.

⁹ Disponible en: <<http://www.youtube.com/watch?v=h0yZkTDTxS4&feature=share>>.

¹⁰ Disponible en: <<https://docs.google.com/file/d/0B9fmLqdAa3HDWVNVncUx5dXVRN2M/edit?usp=sharing>>.



Fig. 1.

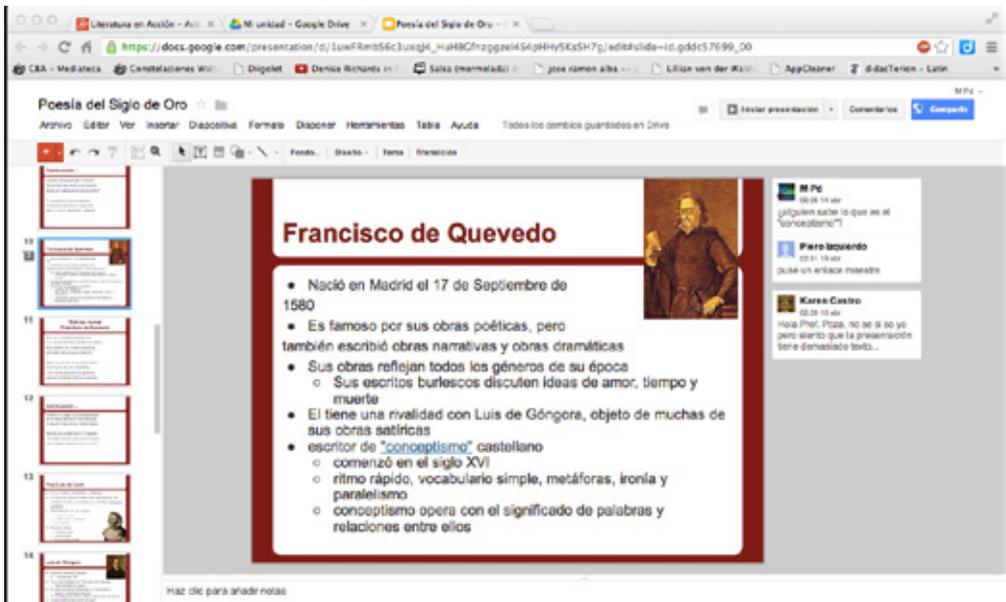


Fig. 2.

contenía su ejercicio escrito. Posteriormente, en el texto *online* se le marcaba mediante colores la palabra, frase o párrafo con el susodicho error, y en una nota al pie se le dejaba el símbolo, de forma que el estudiante debía subsanar la incorrección ortográfica o gramatical. Además, en el lateral del documento, el profesor daba las oportunas indicaciones acerca del contenido del ensayo. Así, se le hacía notar que su tesis no estaba bien formulada, o no resultaba clara, que los argumentos debían conectar con la introducción al trabajo, que las conclusiones debían cerrarse de la manera oportuna. También en los comentarios se le podía sugerir la consulta de tal o cual material, o la revisión del formato para que éste respondiera al estilo propuesto por la MLA. También se le podía conminar a que descartara fuentes inapropiadas para un ensayo académico en beneficio de otras fuentes más rigurosas o serias. Todo ello en tiempo real y desde localizaciones diferentes¹¹.

En este curso todo lo compartido, todo lo producido se integraba en una red social, escogida por sus enormes posibilidades como nuestro CMS. Entre todas ellas, optamos por Google+¹², donde abrimos la Comunidad. Allí se utilizaba el chat, el hangout (multivideoconferencia), YouTube, el paquete de docs, Picassa. Y organizando todo por categorías, compartimos debates que expandieron la clase más allá de las horas lectivas, así como las presentaciones, artículos y materiales. También abrimos allí un foro de ayuda para realizar el *webpaper* e intercambiar ideas. Asimismo, al ser una red social, se pudieron cubrir eventos (como la presentación del libro de Andrés Neuman *Alguien al otro lado*, incluyendo para ello fotografías, enlaces, imágenes, videos, que complementaban la presentación, así como la grabación misma del evento) o la entrevista a través del chat, que los estudiantes también realizaron, al joven autor Juan Gómez Bárcena, al tiempo que se compartían breves análisis de su libro *Los que duermen*, en la página principal de la Comunidad. Allí también, organizados en sus categorías, quedaron pequeñas actividades destinadas a que los estudiantes se familiarizaran con la dinámica pedagógica, pero también con el uso de la tecnología, con propósitos académicos, como los test de lectura, orientados a comprobar si el estudiante había leído el libro correspondiente, o la hoja de cálculo en la que los estudiantes inscribieron sus *webpapers*, como parte de la primera fase de las tres exigidas para su elaboración. Estas actividades corrían parejas al desarrollo del proyecto final: el *webpaper*, que requiere de tres fases de gestación, cada una de ellas susceptible de ser evaluada conforme a determinados criterios que atienden a su desarrollo, a su evolución y al proyecto mismo como resultado final.

La fase uno del *webpaper* conectaba por completo con el examen de *Mid-Term*, ya que el estudiante debía buscar un autor/res, una obra/s, un tema para analizar o investigar, formular una tesis, tres argumentos críticos y elaborar unas conclusiones. En realidad, no hay nada de nuevo en este proceder, se trata de trabajar la más tradicional de las estructuras para crear un ensayo crítico convencional. Aunque, por lo general, los estudiantes se muestran aterrados por tener que realizar el trabajo "final" digitalmente, en realidad, esta clásica fase

¹¹ Video en que puede verse el procedimiento de trabajo en el examen de *Mid-Term*: <<http://youtu.be/xaejMj6BOJA>>.

¹² Problemas técnicos han hecho que gran parte de la información alojada en la Comunidad se perdiera, a pesar de lo cual continúa funcionando. Puede visitarse en: <<https://plus.google.com/communities/107211461125944380755>>. Adjunto imagen de la Comunidad antes de la fatídica pérdida de datos en las figuras 3 y 4.

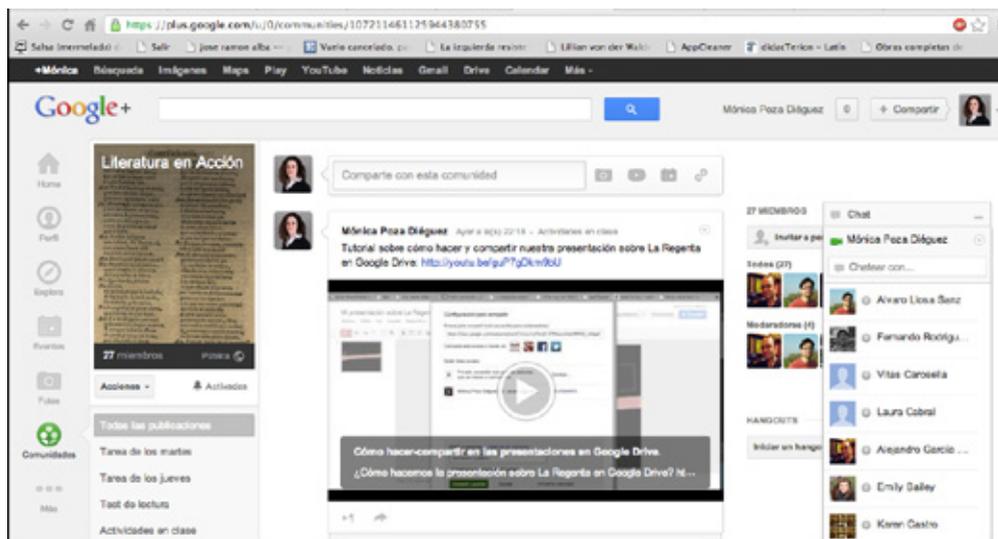


Fig. 3.

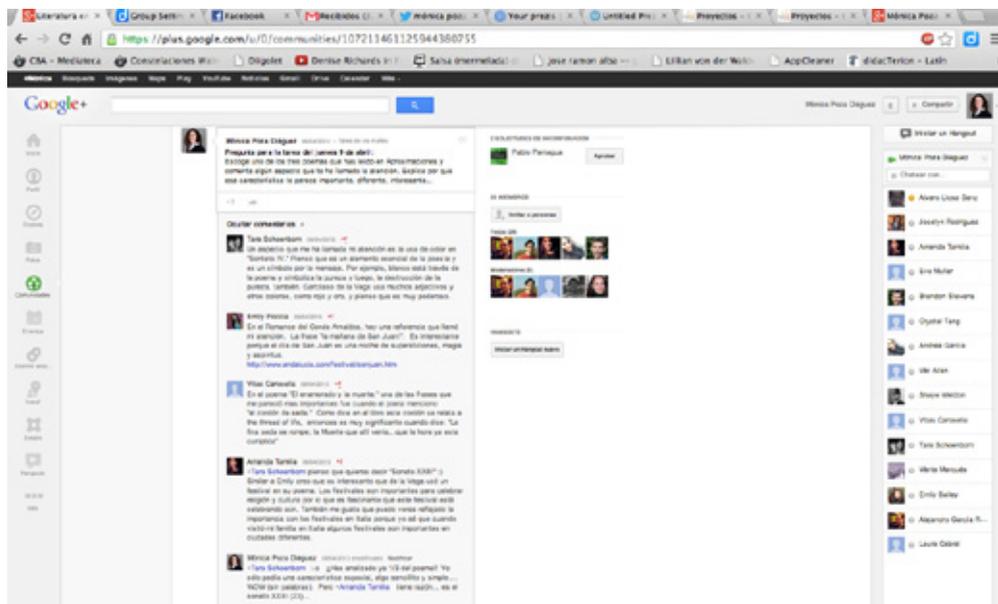


Fig. 4.

para elaborar el ensayo es lo que más les cuesta. Por ello, se les da una fecha para concluir la fase 1, y se les indica que deben completar una hoja de cálculo¹³ en la que deben incluir su nombre y apellidos y todos los campos referentes al ensayo, ya mencionados aquí.

La fase uno se califica con 0 ó 100 puntos, dependiendo de si se completó satisfactoriamente la hoja de cálculo con todos los datos referentes a su proyecto o no. La fase 2 se solapa a menudo con la fase 3, por lo que las fechas pueden desplazarse a conveniencia de todos, aunque es bueno presionar al estudiante para que definitivamente tenga listo el proyecto para la fase 3 a tiempo. La fase 2 consiste en reunir todos los materiales para la investigación y su elaboración. El estudiante buscará y seleccionará desde libros, hasta artículos, imágenes, enlaces, videos, música, o cualquier otro elemento media que conecte con el análisis crítico. Sería ideal que el estudiante elaborara algunos de esos materiales, si existe el tiempo suficiente. Esta fase también se califica con 0 ó 100 puntos, dependiendo de si el estudiante completa o no satisfactoriamente la fase. En la fase 3 se hacen las revisiones conforme a los criterios de evaluación¹⁴, y se calificará de 0 a 100 puntos, conforme a esos mismos criterios y a la calidad general del proyecto.

Estos criterios deben medir la ortografía, la gramática, el contenido, si los materiales empleados conectan bien con los argumentos defendidos, pero también si todo está bien organizado, si comunica satisfactoriamente, si su diseño es agradable. Para la cuestión del diseño se les suele pedir que observen periódicos españoles y americanos, para captar bien la disposición organizada del texto y de las ideas, en el espacio virtual (una cuestión de simple *dispositio* absolutamente imprescindible). Por supuesto, todos los materiales deben estar registrados bajo licencia de Creative Commons (no se pueden aceptar bajo ningún concepto trabajos que incluyan materiales con Copyright). Y todo debe estar citado y referenciado conforme al estilo del MLA en una *webgrafía*. También se debe comprobar que todos los enlaces funcionan correctamente. En la evaluación debe tenerse en cuenta la evolución que el proyecto tuvo a lo largo del proceso de elaboración.

Para los proyectos finales o webpapers¹⁵ The Littera Project ha trabajado sobre todo con Weebly, en su versión educativa, Weebly Education. Se trata de una plataforma muy fácil de usar y muy intuitiva para hacer sitios web. En

¹³ Hoja de cálculo para la fase 1 de los proyectos del curso de iniciación a las literaturas hispánicas Literatura en Acción, impartido en Syracuse University en 2013: <<https://docs.google.com/spreadsheets/ccc?key=0AtfmLqdAa3HDdGVfRkt3TFI2eGhtUkVXUFVKZIBsYWc&usp=sharing>>.

¹⁴ Criterios de evaluación de los proyectos: <<https://docs.google.com/file/d/0B9fmLqdAa3HDUWNfcGpwS2NQcFk/edit?usp=sharing>>.

¹⁵ Ejemplos de algunos webpapers en Syracuse University (nivel subgraduado):

<<http://corintiosysanmanuelbueno.weebly.com/>>.
<[Http://poesiasalvadorena.weebly.com/index.html](http://poesiasalvadorena.weebly.com/index.html)>.
<<http://lamujeryfranco.weebly.com/index.html>>.

Ejemplos de algunos webpapers en la Universidad de Szeged (nivel graduado)

<<http://carnavaldeflores.weebly.com/>>.
<<http://loquedijoelmayordomo.weebly.com/>>.
<<http://elhombrequesaliaporlasnoches.weebly.com/>>.
<<http://pecadodeomision.weebly.com/>>.

Para ver más ejemplos de proyectos, remito a la página web de The Littera Project:

<<http://www.thelitteraproject.weebly.com>>.

este espacio virtual el estudiante debe crear su sitio web conforme a lo establecido en la hoja de cálculo de la fase 1 y con los materiales recopilados en la fase 2. Para ello debe seguir unos criterios retóricos y visuales muy concretos. Cada página o pestaña del website debe corresponderse con las partes de su ensayo. Y todo debe estar claramente organizado a la par que resultar visualmente atractivo.

Para el *feedback* se trabajó con Diigo¹⁶. De este modo, al estudiante se le da un primer *feedback* de *highlights* o subrayado, que aparecerán en su navegador, y cuando visite su propio proyecto. Se marca el párrafo, la frase que debe cambiar o corregir, y en las sticky notes o post-it se le indica el error gramatical u ortográfico que debe subsanar siguiendo la clave de errores ya mencionada. En un segundo *feedback* se le asesorará en cuanto al contenido y los materiales, siguiendo para ello los criterios de evaluación.

Por supuesto, todos los proyectos están publicados, si bien no todos están difundidos, al no reunir todas y cada una de las características que se pedían, pero puede concluirse que, en general, el proceso de aprendizaje por medio de la realización de proyectos y actividades en evaluación continua ha dado buenos frutos, en cuanto a las calificaciones y la motivación que mostraron los alumnos que, cubriendo satisfactoriamente los contenidos del programa, mejoraron enormemente sus habilidades de escritura y orales, asimilaron correctamente el uso de la tecnología y expandieron la clase ampliando sus capacidades críticas a la hora de enfrentarse a un texto.

Las evaluaciones que realizaron de forma anónima sobre el curso fueron muy positivas. Sorprende leer afirmaciones en que se destaca el gran volumen de trabajo al que los estudiantes fueron sometidos al tiempo que se destaca el disfrute, y la convicción de que este "proyecto piloto de la universidad" se expandirá a otras clases.

En conclusión, y volviendo a las preguntas iniciales acerca de si es posible expandir el trabajo académico o *paper* al formato digital. La respuesta es rotunda, no sólo se puede, sino que debe hacerse. El formato digital no resta ni aumenta la calidad del producto académico, pero el cambio pedagógico que conlleva la superación de un modelo institucional académico, basado exclusivamente en lo analógico, para incorporar híbridamente posibilidades digitales, nos da la clave exacta para construir un nuevo modo de comunicar nuestro conocimiento a los estudiantes más acorde con la época en que vivimos. Es necesario incorporar las herramientas tecnológicas en las clases, como se incorporó en su día el libro impreso. No es una cuestión que se nos avecine rápidamente y al modo de un apocalipsis educativo, no. El cambio de paradigma no es una hecatombe, como tampoco proviene del futuro. La crisis paradigmática es una oportunidad y ya está aquí.

BIBLIOGRAFÍA

Abaitua, Joseba, "A ciencia abierta: iniciativas internacionales que catapultan el conocimiento", en *Translema. Joseba Abaitua's Blog on Language and Translation*, <<http://blogs.deusto.es/abaitua/a-ciencia-abierta-iniciativas-internacionales-que-catapultan-el-conocimiento/>> [06/10/2013].

¹⁶ Tutorial-Video *screencast* en que se explica cómo dar el *feedback* con Diigo: <<https://www.facebook.com/photo.php?v=509700525745529&l=702771890003207406>>.

Mattern, Shannon Christine, "Evaluating Multimodal Work, Revisited", en *Words in Space*, <<http://www.wordsinspace.net/wordpress/2012/08/28/evaluating-multimodal-work-revisited/>> [09/10/2012].

Reig, Dolors, Luis F. Vílchez, *Los jóvenes en la era de la hiperconectividad*, <http://www.fundacion.telefonica.com/es/arte_cultura/publicaciones/add_descargas?doc=Los%20jóvenes%20en%20la%20era%20de%20la%20hiperconectividad:%20tendencias,%20claves%20y%20miradas&pdf=media/publicaciones/los_jovenes_en_la_era_de_la_hiperconectividad1.pdf&type=publicacion&code=182> [15/07/2013].

Salido Rojo, María Jesús, *Comunidades de práctica. Una metodología para construir, desarrollar, y fortalecer redes de conocimiento*, <<http://www.bubok.es/libros/212443/Comunidades-de-Practica-Una-Metodologia-para-Desarrollar-Construir-y-Fortalecer-Redes-de-Conocimiento>> [26/01/2013].

Smith, Sidonie, "Beyond The Dissertation Monograph", en Modern Language Association (ed.), *MLA Newsletter*, <<http://www.mla.org/blog&topic=133>> [09/09/2010].



RESUMEN

En *The Littera Project* un grupo de profesores desarrollan cursos *online* colaborativos, multimodales y transmedia, donde los estudiantes trabajan en proyectos de investigación. En este artículo muestro cómo se puede expandir el formato de la monografía tradicional para desarrollar un *webpaper* (un trabajo académico digital).

Palabras clave: The Littera Project, educación superior digital, transmedia, multimodalidad, trabajo académico digital.

ABSTRACT

In *The Littera Project* a group of professors are developing online collaborative multimodal and transmedia courses, where students work on research-based projects. In this article I show how we can expand the format of the traditional monograph to develop a *webpaper*.

Keywords: The Littera Project, digital higher education, transmedia, multimodality, *webpaper*.

TEI como una nueva práctica de lectura

ERNESTO PRIANI SAISÓ
Universidad Nacional Autónoma de México
epriani@gmail.com

ANA MARÍA GUZMÁN OLMOS
Universidad Nacional Autónoma de México
sositap@gmail.com

Alguien lee un libro. Mientras lo hace, subraya. También pega pequeñas tiras de colores para hacer más fácil la identificación de páginas y pasajes que podrían ser útiles en el futuro. Escribe en los márgenes notas con juicios e ideas que la lectura le produce. De esta forma, con paciencia, mientras se lee, el texto en el libro se va llenando de señales. Un día ese libro es prestado. Pasa a manos de otro. Alguien que tiene una edad, un punto de vista, una historia diferente a la de aquel que ha dejado las marcas sobre el texto. Es un otro que lee la misma obra pero con otras necesidades, con otros intereses, buscando cosas distintas.

Las señales en el texto que materializan una lectura, representan para el nuevo lector un estorbo, pero también un diálogo. Estas son registro de las cosas que habían resultado interesantes a otra persona. Evidencias que dificultan la lectura porque se interponen entre el nuevo lector y el texto, sesgándolo, distrayéndolo de sus propósitos, de sus intereses. Pero estas mismas señales abren otra posibilidad: la de reconstruir esa primera lectura. A fin de cuentas, marcar significa dejar las huellas de los pasos andados en un texto. Pistas para reconstruir el intangible acto de leer.

El momento de subrayar y señalar un texto nos permite reconocer en la lectura los rasgos propios de una práctica cultural. Pues, como señala Pierre Mayol, una práctica "es el conjunto más o menos coherente, más o menos fluido, de elementos cotidianos concretos (un menú gastronómico) o ideológicos (religiosos, políticos), a la vez dados por una tradición (la de una familia, la de un grupo social) y puestos al día mediante comportamientos que traducen en una visibilidad social fragmentos de esta distribución cultural, de la misma manera que la enunciación traduce en el habla fragmentos de discurso" (2010: 7). Al materializar la lectura se pone en evidencia que se encuentra compuesta por un conjunto de hábitos y costumbres que dan sentido, en un determinado momento y lugar, al acto de leer.

Subrayar un texto forma parte del conjunto de actividades que componen la lectura, a través de hábitos como el de ir dejando señalamientos se da visibili-

dad a esta práctica. Si, como indica Chartier, "la lectura no es solamente una operación intelectual abstracta: es una puesta a prueba del cuerpo, la inscripción en un espacio, la relación consigo mismo o con los demás" (2011), entonces podemos afirmar que la lectura es una práctica en tanto se inscribe en un espacio determinado (se lee en bibliotecas, cafés, parques, etc.), se relaciona con un conjunto de objetos (textos), se lleva a cabo por ciertos sujetos (lectores) en relación con otros (editores, escritores, vendedores), está rodeada por un grupo de instituciones que conforman una tradición de lectores, entre otros factores que dan sentido a la lectura.

Ahora, pensemos que el libro marcado y señalado no está contenido en la objetualidad de un volumen. Es decir, no es ya un libro impreso con papel y tinta. Pensemos en el texto digital. Aún cuando lo que aparece en la pantalla son también letras, tal y como aquellas de la imprenta, las operaciones que han sucedido para que el texto se despliegue en ella no nos permiten decir que se trata exactamente del mismo. Algo ha tenido que ocurrirle al texto para que podamos verlo ahí. Pues no podemos hacer con él (al menos no del mismo modo) lo que normalmente haríamos con un impreso. No dejamos la huella de nuestra lectura de forma idéntica de cuando lo hacemos sobre el objeto.

Que un texto sea digital significa, de manera general, que lo hemos convertido en datos pues, desde el punto de vista del cómputo, el texto es "información codificada como caracteres o secuencia de caracteres", con independencia de cualquier significado que estos tengan. En este sentido, como han escrito Buzzetti y McGann: "la forma digital del texto lo define como un objeto sobre el cual las computadoras pueden operar algorítmicamente, para cargarlo de sentido e información" (2013).

Este profundo cambio en el texto no lo afecta únicamente a él. En realidad implica una serie de modificaciones que trastocan todo lo que lo rodea. Desde el lugar donde se despliega, la manera en que puede ser manipulado, el grupo de personas que tienen contacto con él (pues no necesariamente son las mismas que intervienen en la producción, difusión y lectura de un libro) y la forma en que estos, particularmente el lector, se relaciona con el texto y el modo en que dejan huella de su lectura.

"Al romper el antiguo lazo anudado entre los textos y los objetos, entre los discursos y su materialidad, la revolución digital obliga a una radical revisión de los gestos que asociamos con lo escrito. A pesar de la inercia del vocabulario que intenta domesticar la novedad denominándola con palabras familiares, los fragmentos de textos que aparecen en la pantalla no son páginas, sino composiciones singulares y efímeras" (Chartier, 2011: 21).

Como señala Chartier, ante la necesidad de referirnos al texto digital de algún modo, recurrimos a palabras que ya conocemos y que provienen del mundo del libro. Hablamos de páginas, de capítulos, de volúmenes, de tipografía, de subrayados... Es a través de este curioso fenómeno de llamar a cosas nuevas con los nombres de antes, que podemos atestiguar cómo se está produciendo la transición de una práctica con la que estamos familiarizados hacia una emergente a la que, en realidad, no conocemos. En nuestra experiencia, lo que notamos cada vez más es una inadecuación entre el modo con que llamamos a lo que hacemos y lo que en realidad estamos haciendo con el texto digital. Pues la forma con que nos relacionamos con él, lo alteramos, lo manipulamos,

lo leemos y lo señalizamos, nos obliga a replantearnos la relación entre el texto, las prácticas de sus lectores y los conceptos con los que nos referimos a ellas.

Para discutir estas relaciones, partiremos del trabajo que hemos llevado a cabo con los textos que forman la Biblioteca Digital del Pensamiento Novohispano y de la experiencia de transformar estos textos desde su condición objetiva hacia su digitalización. Nos centraremos, sin embargo, en el hecho de que durante el proceso de digitalización se produce el registro de una lectura y en cómo, a partir de este hecho, se desarrolla una nueva práctica de lectura, muy diferente a la que se hace con el papel.

1. MARCANDO EL TEXTO

La Biblioteca Digital del Pensamiento Novohispano es una iniciativa de la Facultad de Filosofía y Letras de la UNAM. Tiene su origen en la intención de editar digitalmente ocho documentos que registran una disputa sobre la naturaleza de los cometas que tiene lugar en la Nueva España durante la segunda mitad del siglo XVII, en la que una parte de los contendientes sostiene que los cometas son de origen terrestre, mientras otros argumentan que se trata de cuerpos supralunares, y por lo tanto, celestes. Con esta visión primaria, en su concepción original, el proyecto únicamente buscaba hacer una edición digital, en la que los textos estuvieran acompañados de algunos índices que pudieran ser útiles para su lectura.

Después de un intento por generar un esquema de marcado XML, se tomó la decisión de usar el estándar de marcado de las guías TEI cuyos parámetros en la definición de sus elementos y atributos permitía al menos dos cosas importantes. Por un lado, era posible tener un esquema flexible que permitiera cambiar la definición de las marcas, eliminar o introducir nuevas sin tener que redefinir la totalidad del esquema. Por otro lado, permitía tener a disposición una serie de *tags* cuya coincidencia con nuestra idea de la edición de texto era muy amplia. Una vez que nos convencimos de usar TEI, definimos la serie de *tags* que serían útiles tanto para representar digitalmente los textos, como para marcar las palabras que queríamos recuperar como índices. Así, comenzamos utilizando cinco marcas: `<name>`, `<term>`, `<date>`, `<q>`, `<cit>`, `<refTitle>`, `<foreign>` para identificar *nombres, términos, fechas, citas, referencias, títulos referidos y textos en otro idioma*. A algunas de estas les añadimos algunos atributos porque buscábamos recuperarlas con alguna especificidad. Así, por ejemplo, `<name>` adquirió los atributos: *person, place, planet, constellation, god, y zodiac*, `<term>` los atributos `<med>` médico, `<astro>` astrológico, `<astron>` astronómico, `<phil>` filosófico, `<math>` matemático, `<theo>` teológico y `<oldSpa>` español antiguo, mientras `<date>`, el de `<comet>` cometa y `<foreign>` el de `<lat>` latín.

Es curioso ver cuánto tiempo tardamos en darnos cuenta que al marcar el texto con estas etiquetas estábamos haciendo algo más que identificarlas para formar un índice. No teníamos entonces, y no tenemos aún, un concepto (fuera del de marcar) que identifique claramente qué es lo que se hace cuando uno comienza a preguntarse cosas como si la palabra "fuentes" puede ser un término médico, o si los nombres de lugares son lugares, o cómo clasificar un marquesado. Porque la marca no sólo es un identificador, es un dispositivo hermenéutico que sirve para construir el significado de una palabra.

Esto lo observamos al notar que las palabras no se integraban de manera intuitiva al conjunto definido por alguna de las etiquetas, sino que necesitaban ser interpretadas y que esa interpretación se hace en función de las propias etiquetas con las que se cuenta. Es decir, como buscábamos términos médicos dentro de un lunario del siglo XVII, encontramos que "fuentes" era utilizada como referencia a un procedimiento médico que podía practicarse en ciertos días del mes. Cuando uno marca, la lectura del texto está condicionada por la posibilidad de interpretación que ofrece la etiqueta.

Al abordar el texto de la *Libra Astronómica y Filosófica* de Carlos de Sigüenza y Góngora, nos encontramos con numerosas menciones a marquesados y marqueses. La frecuencia con que se hacía nos llevó a pensar en la conveniencia de marcarlos, pero siempre sobre la base de hacerlo bajo una de las etiquetas previamente definidas. En este caso el problema era que, por una parte, se hacía referencia a una persona específica: El Marqués de Cáceres y, por otra, el nombre del título nobiliario, que usualmente está vinculado con un lugar; por ejemplo el Marquesado de Cenete. Qué debíamos marcar: ¿lugares?, ¿instituciones? ¿nombres propios? Cualquiera que fuera la decisión interpretativa que tomáramos, la lectura estaba mediada por las etiquetas que podíamos utilizar.

En efecto, el uso de marcas en el texto, como ha señalado Fiormonte, no es neutral. Supone siempre una toma de posición aun cuando esta no haya sido declarada explícitamente y no corresponda a un marco teórico preestablecido, como ocurre por ejemplo en el caso de los marquesados. Una marca dice, de hecho, algo respecto del objeto marcado, como es muy claro en el ejemplo de "fuentes". Es una interpretación, y supone una serie de operaciones teóricas respecto de lo marcado, que no son triviales. Pero además, puesto que esas marcas son procesadas computacionalmente ofrecen un resultado visible –por ejemplo en nuestro índice– en que se confirman o se ponen en cuestión los supuestos y las operaciones teóricas llevadas a cabo para marcar.

Un ejemplo de esto último fue la decisión de dejar de utilizar *<name type = "place">* para definir hacer un índice de lugares, cuando decidimos geolocalizarlos. Entonces nos dimos cuenta de que los nombres no tienen ubicación geográfica, es decir, que dentro del uso del etiquetado de TEI la geolocalización es un atributo posible de *<place>* y no de *<name>*. Así que, si queríamos efectuar una cierta operación computacional, recuperar una palabra no en un índice sino en un mapa, teníamos que resignificarla por completo.

Pero lo más significativo de colocar una etiqueta en una palabra es que esa operación interpretativa queda registrada y es puesta a prueba con el procesamiento del texto.

De cierta forma, con el marcado, hacemos algo muy parecido a lo que hacía la persona de la que hablábamos al principio: marcar significa hacer ese procedimiento de selección y de decisión sobre lo que sería deseable encontrar en un texto sin tener que leerlo todo detenidamente. Pero, a diferencia de marcar poniendo una línea o un color, lo que hacemos al usar TEI es utilizar palabras para etiquetar. Las palabras con las que marcamos tienen un sentido, no son formas o colores, sino conceptos que agrupan los objetos marcados adhiriéndose a ellos. Poner *<name>* a un objeto que también es una palabra, significa el surgimiento de un nuevo sentido que acontece en esa unión. Esta característica singular del marcado es lo que ha hecho reflexionar a Singer sobre su aplicación como técnica de lectura didáctica para estudiantes, pues ante todo, permite re-

flexionar sobre los diferentes vocabularios que utilizamos para describir un texto y, agregaríamos, las implicaciones hermenéuticas que cada uno tiene.

“While many digital sites were originally built for scholars, it seemed time to consider what kinds of coding and interfaces might work best for students. Most importantly, in a course designed to think carefully about poetic literary terms, TEI seemed like a wonderful way to teach hands-on close reading. Perhaps one of the most interesting and important skills TEI can teach is to help our students become reflective about the various kinds of vocabularies they use to describe texts” (Singer, 2012).

Singer ha propuesto, después de observar a sus estudiantes marcar un texto, que leer para marcar es diferente a la forma en que nosotros leemos usualmente porque la estructura de la lectura es claramente distinta. Cuando usamos etiquetas se produce una reflexión sobre la forma en que las marcas describen las palabras y la manera en que la etiqueta establece su sentido. Algo que no ocurre con un subrayado cuya función suele ser mucho más denotativa que interpretativa. Además, el sentido establecido para la palabra se conserva aún después del procesamiento de lo marcado, lo que significa que la interpretación se manifiesta a otros lectores. Cuando leemos para marcar, vamos al texto guiados por una búsqueda de palabras/frases marcables.

Pero hay algo más en este proceso de lectura para el marcado. Normalmente, aquel que subraya un libro no lo hace pensando en que sus señalizaciones vayan a ser “leídos” por otro. Los subrayados son considerados marcas personales, individuales, que se vinculan con la propiedad de un libro. Dentro del código de conducta de la lectura está, por supuesto, no devolver un libro prestado con un subrayado, a un amigo o a una biblioteca. Las marcas en un libro de hecho lo degradan ante los ojos de los demás. Pero en el caso del marcado de texto, las marcas se hacen pensando en que el texto marcado será, por un lado, procesado y su resultado puesto a la mirada de otros, convirtiéndolas en un elemento integral de la lectura digital.

¿Cómo cambia la figura del lector con la búsqueda de objetos marcables? El lector de un texto impreso se identifica con individuos que eventualmente se encuentran frente al libro, pero la figura del lector en el marcado digital está integrado tanto por individuos como por el modo en que se procesa el texto. La máquina produce acciones vinculadas al texto, la figura del lector está compuesta también por esas determinaciones. No es entonces una alteración menor la que acontece en la lectura del texto que se marca, hay un nuevo elemento que constituye el marco de interpretación ante el que se llena de sentido al texto. La computadora se convierte en un lector.

De este modo, etiquetar digitalmente un texto implica considerar que esas marcas van a integrarse a la lectura que otro haga del texto. Agregamos, pues, a la lectura de los demás, la naturaleza de la lectura que hagamos para marcar. Lo que implica una conciencia claramente diferente del lector común que subraya un libro. Pues al incorporar una lectura al cuerpo del texto que se lee para mostrarlo digitalmente, alteramos por completo el proceso de la lectura, pues se lee un texto que ha sido previamente leído de un modo, que no es además el modo usual en que se leen los textos. En este sentido, no podemos sino coincidir con Chartier: “las mutaciones de nuestro presente modifican todo a la vez, los soportes de la escritura, la técnica de su reproducción y diseminación y las maneras de leer” (2011).

Marcar textos es, pues, una forma inédita de leer, guiada por la existencia del dispositivo hermenéutico de las etiquetas y la deliberada búsqueda de palabras/frases que pueden ser marcadas. Esta lectura a su vez, produce un texto distinto para ser leído: el que incorpora como parte propia las etiquetas que condicionan su representación en pantalla.

2. IR Y VOLVER: LEER PARA INVESTIGAR

Hemos dicho ya, que las marcas materializan la lectura. Constituyen una evidencia de un proceso hermenéutico, de una forma de interpretación de los textos, que pasa a formar parte del propio texto procesado. Y esto permite una lectura a la vez del texto y de sus marcas. En otras palabras, el etiquetado puede ser analizado y estudiado una vez que ha sido procesado.

En el contexto de nuestro trabajo en la BDPN hemos observado cómo, una vez procesados los textos se revelan dificultades e inconsistencias en la interpretación. Lo mismo que omisiones y equivocaciones, pero también y esto es lo más importante, la posibilidad de encontrar algo en el texto que no habíamos visto al inicio. Es decir, la posibilidad de volver a leer y obtener conocimiento nuevo. Este proceso de ir y volver sobre el texto, de guiarnos por las marcas para identificar algo que había pasado desapercibido, es quizás la mayor aportación de esta nueva forma de leer.

Uno de los aspectos de la investigación que desarrollamos en la Biblioteca Digital del Pensamiento Novohispano, es la identificación de las fuentes de la disputa astrológica, con la intención de valorar el grado de influencia del pensamiento bizantino, medieval, renacentista y moderno en la construcción del discurso sobre los cometas. La estrategia utilizada fue marcar nombres, citas, referencias y paráfrasis, para posteriormente ir identificando en cada caso de quién y de qué obras se trata. Uno de los nombres marcados tratándose de una disputa sobre la naturaleza de los cometas fue, por supuesto, Ptolomeo, así como las obras que le son atribuidas en los distintos documentos que marcamos.

Nuestra expectativa en relación con Ptolomeo todo el tiempo fue que sería referido como una autoridad y defendido en aquellos documentos donde prevalece la idea de que los cometas son supralunares y, por tanto, producto de las exhalaciones terrestres. A fin de cuentas, Ptolomeo era partidario de esa idea y de manera natural sería utilizado por quienes la defienden. Nuestra sorpresa fue descubrir, una vez procesados los textos, que Ptolomeo era recurrido como autoridad por ambos bandos, pero no de la misma manera o al menos no refiriendo las mismas obras.

Para los detractores de la idea de la naturaleza celeste de los cometas, la fuente de referencia como autoridad es Ptolomeo en el *Quadripartito o Tetrabiblos*. Así, por ejemplo, Joseph de Escobar Salmerón y Castro, y Martín de la Torre, polemistas y opositores a Sigüenza, refieren como autoridad a Ptolomeo como autor de una obra de astrología judiciaria. Por ejemplo, de la Torre:

“Yndica segun Ptolomeo Muerte y [Ynterivo] de Principes y Cavezas Coronadas, y porque ese mismo movimiento fue hacia el Orientee, amenaza guerras estrangeras invasion y asaltos repentinos, y mas en las regiones y parte que ha sido perpendicular en el Zenit, o en cuyos Horizontes fué mas larga su demora y aparicion, quales son todas las de poca altura de la una y otra [vanda] de la linea; porque en el principio tuvo este Cometa con poca diferencia la misma declinacion de la

vanda del sur que la espiga de la Virgen que es de [9] grados y [25] minutos aunque esta siempre fué amenos, hasta que el cometa pasó la linea, con declinacion Septentrional, que va creciendo de día en día" (BDPN 2013).

Y de la misma forma Salmerón y Castro refiere el Centiloquio diciendo "Que ambas a dos Estrellas (refiriéndose al Sol y a Marte) son de horrible, y funebre Naturaleza, y significacion, como consta del Centiloquio 73. de Ptolomeo, y de sus Expositores." Pero no pasa lo mismo con Sigüenza. Este escribe, sin rebatir la autoridad de Ptolomeo,

"O Santo cielo! Es posible que Claudio Ptolomeo Autor del *Almagesto*, Principe de la Astronomia, aquel que en cap. I desta grande obra afirma haberse aplicado al estudio y especulación de las Mathematicas por la indefectibilidad destas ciencias, y no al de la Theologia y Philosophia, por tener una y otra por fundamento las congruencias y congeturas ... Es posible vuelvo a decir que este dedicase su vigilancia y consumiese el preciosísimo tempro de mucha horas en escrevir de la Astrologia, cosa que carece de fundamento, de reglas scientificas, de acolutia!" (BDPN 2013).

Para Sigüenza y Góngora, Ptolomeo es el autor del *Almagesto*, pero no del *Quadripartito* o *Tetrabiblos* o del *Centiloquio*, obras de astrología judiciaria. Si entendemos bien, la cuestión no es negar la autoridad de Ptolomeo como científico y padre de la astronomía, sino aceptar que haya escrito un texto de astrología judiciaria, siendo que esta no puede ser considerada científica. De modo que las obras que justifican la posibilidad de que los astros tengan una influencia sobre el mundo sublunar, no podría estar escrito por el fundador de la astrología propiamente científica.

Así, observamos que los partícipes en la disputa del siglo XVII no concuerdan en atribuir a Ptolomeo los textos de astrología predictiva, y en cambio coinciden en atribuirle los textos que consideraban de contenido científico. De manera que ninguno discute la autoridad de Ptolomeo, pero sí los campos sobre los cuales se ejerce su autoridad.

Este tipo de conclusiones que han sido obtenidas una vez que el texto ha sido procesado, permiten observar de qué modo el registro material de una lectura, da lugar a una lectura de segundo orden. Tanto índices como visualizaciones, permite identificar las hipótesis sobre las cuales un texto ha sido leído, y observar qué tanto se ajustan a esa hipótesis inicial, para después proponer o formular otras hipótesis. Tal como nos ha ocurrido con Ptolomeo.

Resumiendo lo que hemos expuesto hasta aquí, el marcado de texto da lugar a un proceso de lectura compuesto al menos de dos momentos. Uno inicial, en que a partir de etiquetas determinadas se marcan palabras. Es una suerte de camino de ida, donde las hipótesis iniciales son reflejadas en el cuerpo de texto. Y un camino de vuelta, que consiste en leer el texto a partir de las marcas dejadas sobre él. En estos dos momentos, cada uno con su propia complejidad y sus propias implicaciones hermenéuticas, descubrimos una práctica de lectura que se genera en y desde el texto digital. Y donde, como afirma Ciccoricco, se produce una poderosa relectura:

It follows then that close reading digital literature is inevitably close re-reading. Granted, close reading of any literary text involves re-reading, but narrative and poetic texts in digital environments take the practice of re-reading to a higher power or second-order. Such texts, because of their changeability, recursivity, and multi-linearity "rely on reiteration for their iteration; that is, re-reading can

no longer be thought of as an epiphenomenon of reading in a network text since the re-reading of textual elements, via the recycling of nodes, is fundamental to (hyper)textual comprehension" (2012).

¿Por qué a todo esto le damos el nombre de una nueva práctica de lectura? Básicamente en tanto se modifica la relación del lector con el texto. Esta forma de lectura digital –por darle algún nombre– está mediada por el uso de las marcas. Las etiquetas guían, establecen dimensiones y campos de la lectura, y la constriñen. Luego, median también en la construcción de la re-lectura porque alteran el texto al quedar incrustadas. Hacen pues factible una nueva lectura con ellas. Esta práctica de lectura requiere pues de herramientas y procedimientos distintos a los que hacemos en papel. Requiere definición y estructura para las marcas, la formulación de hipótesis explícitas respecto del texto. Después, la intermediación de un algoritmo, con el que se procesan las marcas y un despliegue visual donde esas marcas se hacen manifiestas. Un ejercicio, pues, con instrumentos, métodos y fines, que constituyen otra forma de leer.

"Por consiguiente, una historia de largo alcance de las lecturas y los lectores ha de ser la de la historicidad de los modos de utilización, de comprensión y de apropiación de los textos. Considera al 'mundo del texto' como un mundo de objetos, formas y ritos cuyas convenciones y disposiciones sirven de soporte y obligan a la construcción del sentido. Por otro lado, considera asimismo que el 'mundo del lector' está constituido por 'comunidades de interpretación' (según la expresión de Stanley Fish), a las que pertenecen los lectores/as singulares. Cada una de estas comunidades comparte, en su relación con lo escrito, un mismo conjunto de competencias, usos, códigos e intereses" (Chartier, 2011: 27).

Hay una serie de hábitos vinculados al uso de TEI que van construyendo poco a poco nuevas comunidades de interpretación y con ello nuevos códigos, nuevas formas de pensar al texto. No es pues accidental la manera en que se trabaja con el texto a través del marcado, se trata de un factor determinante para la forma de comprender lo que se lee, pues hay un conjunto de hipótesis que pueden ser formuladas en tanto se plantean en función del uso de marcas (*tags*).

Si una práctica de lectura está vinculada a una cierta materialidad de los modos en que se presenta el texto debe aparecer una modificación cuando lo que se lee no es de hecho material. El cambio sobre la perspectiva con la que se lee, los elementos implicados en la proyección de la figura del lector permiten hablar de una nueva práctica. Pero todos estos elementos generan también nuevas vías de investigación por lo que valdría la pena preguntarse si el uso de TEI no ha llevado, no sólo a leer de manera distinta, sino a generar una nueva forma de investigar, una en la que la figura de alteridad está implicada desde la manera misma en que se guía la mirada a través del texto, hasta la forma en que se da sentido a cada una de las palabras.

BIBLIOGRAFÍA

- Biblioteca Digital de Pensamiento Novohispano, <<http://www.bdpn.unam.mx>> [05/07/2013].
 Buzzetti, Dino y McGann, Jerome, "Electronic textual editing: critical editing in a digital horizon", en *Text Encoding Initiative*, TEI Consortium, <http://www.tei-c.org/About/Archive_new/ETE/Previewmccgann.xml> [05/07/2013].
 Chartier, Robert y Cavallo, Guglielmo (coords.), *Historia de la lectura en el mundo occidental*, México, Taurus, 2011.

- Ciccoricco, David, "The materialities of Close Reading: 1942, 1959, 2009, 2012", en *Digital Humanities Quarterly*, The Alliance of Digital Humanities Organizations, <<http://www.digitalhumanities.org/dhq/vol/6/1/000113/000113.html>> [05/07/2013].
- Fiormonte, D.; Martiradonna, V. y Schmidt, D., "Digital Encoding as a Hermeneutic and Semiotic Act: The Case of Valerio Magrelli", en *Digital Humanities Quarterly*, The Alliance of Digital Humanities Organizations, <<http://www.digitalhumanities.org/dhq/vol/4/1/000082/000082.html>> [05/07/2013].
- Mayol, Pierre, "Habitar", en Michel de Certeau *et al.*, *La invención de lo cotidiano*, México, Universidad Iberoamericana e Instituto Tecnológico de Estudios Superiores de Occidente, 2010.
- Priani, Ernesto, "El marcado de texto como representación hermenéutica", en *Horizontes de interpretación: la hermenéutica y las ciencias humanas*, tomo 1, UNAM, 2012.
- Singer, Kate, "The Melesina Trech Project: Markup Vocabularies, Poetics, and Undergraduate Pedagogy", en *Digital Humanities 2012*, <<http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/the-melesinatrench-project-markup-vocabularies-poetics-and-undergraduate-pedagogy/>> [05/07/2013].



RESUMEN

El propósito de esta comunicación es mostrar cómo el marcado de texto con TEI puede ser entendido como una práctica de lectura diferente e independiente de otras, a través de una serie de ejemplos tomados de nuestra investigación con la Biblioteca Digital del Pensamiento Novohispano (<www.bdpn.unam.mx>).

El uso de algunas marcas (*tags*) particulares, implican decisiones sobre el texto y formulación de una serie de hipótesis acerca de él. Leer para marcar es diferente a la forma en que leemos usualmente porque no buscamos responder a las mismas preguntas. Con la premisa de que lo marcado serán datos para ser procesados con una computadora, marcamos pensando qué se puede hacer con esos objetos.

En el proyecto de la BDPN hemos notado que con el procesamiento de los datos marcados también se generan nuevas lecturas. Esas lecturas las explicamos así: marcamos el texto con ciertas expectativas e hipótesis, pero el procesamiento de lo marcado (índices, visualizaciones) crea nuevos acercamientos al texto y nos permite hacer preguntas que no estaban, necesariamente, implicadas en nuestro primer acercamiento. La producción de índices como resultado del procesamiento de texto, nos permite desarrollar una nueva lectura del texto.

Si un texto se encuentra unido a una práctica de lectura, el texto marcado, modificado por las marcas que usamos, está, necesariamente, vinculado a una práctica de lectura distinta. Describimos esa práctica como un ir y volver en el texto, formada por el proceso de marcado y el procesamiento de los datos.

Palabras clave: TEI, lectura, marcado de texto, biblioteca digital.

ABSTRACT

The purpose of this paper is to show how the markup of text with TEI tags can be understood as a practice of a different reading and autonomous from others with a series of examples taken from our research done with the Digital Library of Novohispanic Thought (<www.bdpn.unam.mx>).

Using some particular tags implies making decisions and formulating hypothesis about the text. Reading in order to markup is different to the way in which we usually read, because we are not looking to answer the same questions. The markup is done with the

premise that what we marked will turn into data, we mark thinking on what could be done with this objects.

In the BDPN project we have noticed that through the processing of the data that has been marked, new readings are generated. These new readings are created as follow: we marked a text with certain expectations and hypothesis, but the processed data (index, visualizations) creates new approaches to the text and allows us to make questions which were not, necessarily, implied in our first approach. The indexes production as a result of the text processing, allow us to develop a new reading of the text.

If a text is linked to a practice of reading, the marked text modified by the tags will be linked to a different practice of reading. We describe that practice as going forward and backward in a text, shaped by the markup process and the data processing.

Keywords: TEI, Reading, Markup, Digital Library.

Innovación y creatividad: enseñar literatura en el siglo XXI

BEGOÑA REGUEIRO SALGADO
Universidad Complutense de Madrid
bregueirosalgado@edu.ucm.es

Siglo XXI. Sociedad digital. Teorías constructivistas. LITERATURA. ¿Aprendizaje o transmisión? ¿Cuál es la literatura del siglo XXI?

Como señala Teresa Colomer (1996), la reflexión sobre la didáctica de la literatura comienza alrededor de los años sesenta del siglo XX, cuando los críticos y los educadores empiezan a darse cuenta de lo obsoleto del sistema educativo implantado en el siglo XIX.

Un primer paso para separarse de ese sistema fue el comentario de texto, que pretendía conseguir una comunicación más directa entre el alumno y la obra, de acuerdo con las teorías estructuralistas y formalistas. Sin embargo, su carácter excesivamente sistematizado, pautado, codificado lingüísticamente y sujeto a evaluación se convertía en una pesadilla, más que en un espacio para el aprendizaje, el crecimiento personal y el placer. Según Benigno Delmiro (2001: 29): "Acaba, así, convertido en un frío esquema con el que tratamos de corroborar en un texto los rasgos de época o movimiento estético apuntados a los alumnos y alumnas a propósito de la teoría".

Resultaba, pues, un nuevo fracaso que marcaba la necesidad de dar un paso más.

Afirma Teresa Colomer: "La enseñanza de la literatura resulta muy sensible a los cambios producidos en los mecanismos de producción cultural y de cohesión social de los distintos momentos históricos" (1996: 123). Si, junto a esta cita, situamos la de Román Gubern, según la cual: "La 'pantallización' de la sociedad moderna¹ es un fenómeno muy característico y constituye una llamativa señal de identidad de nuestra civilización tecnológica" (2010: 94), parece que se va atisbando una conclusión, que queda clara tras entender el concepto de pedagogía de la rutina, del mismo Gubern (2010), y leer la advertencia de Umberto Eco en su diálogo con Jean-Claude Carrière: "La rapidez con la que la tecnología se

¹ El mismo autor habla poco más adelante de la brecha digital, por lo que hemos de precisar que se refiere a nuestra sociedad occidental de primer mundo, o a lo que él llama, parafraseando a Jorge Semprún, "el balneario".

renueva nos obliga a un ritmo insostenible de reorganización continua de nuestras costumbres mentales” (Gubern, 2010: 120). Así pues, sólo nos queda unir la línea de puntos: si vivimos en una sociedad en la que nos rodean las pantallas en todos los ámbitos y en la que la información está a un *clic* de nosotros, hemos de aceptar que los procesos cognitivos están cambiando y, de acuerdo con ello, hemos de transformar también nuestros métodos de enseñanza para convertirlos en procesos de aprendizaje en los que el alumno sea el protagonista y el constructor de su propio conocimiento. Si a esto añadimos la especial vinculación de la literatura a la sociedad, como agente constructor y transmisor del imaginario colectivo y la cultura, es obvio que la forma en la que trabajamos en el aula ha de cambiar necesariamente.

A partir de todo ello, en este trabajo, parto de una constatación y de dos ideas para llegar a un propósito. La constatación es la del carácter obsoleto y caduco del sistema historicista de enseñanza de la literatura, así como del sistema tradicional del comentario de texto. Las dos ideas se refieren a lo siguiente:

1. La literatura está íntimamente relacionada con la sociedad y, tal como afirma Teresa Colomer (1996) la forma de enseñarla dependerá del papel que la sociedad le otorgue.
2. Según las corrientes constructivistas, lo idóneo es que el estudiante cree su propio conocimiento y, por ello, de acuerdo con la idea de pedagogía de lo cotidiano de Gubern y con la pantallización de la sociedad, no podemos obviar las pantallas a la hora de guiar a nuestros alumnos en su aprendizaje.

Así pues, el propósito de este trabajo es ofrecer posibles líneas de actuación que favorezcan una nueva forma de trabajar la literatura en el aula a todos los niveles.

Para ello, es necesario que nos detengamos un momento en una de las ideas apuntadas por Colomer: el papel que la sociedad otorga a la literatura en este momento. Si, por un lado, la reducción de horas de lengua y literatura en los institutos, y el decaimiento del prestigio de las carreras de letras han llevado a muchos a considerar inútil la literatura, en un contexto en el que todo lo que no produzca resultados materiales inmediatos parece prescindible, hay voces que se levantan contra esto y nos recuerdan lo fundamental de la literatura en la formación de la persona y en la transformación del mundo en el que vivimos. Así, Guadalupe Jover (2007) señala que: “Uno de los retos actuales de la educación literaria es conciliar el placer de la lectura y la enseñanza de la literatura, conseguir que esta última sirva para ensanchar –y ahondar– las posibilidades de disfrute de los textos literarios. Tanto como anclarnos en una determinada tradición cultural, la enseñanza de la literatura debiera desarrollar nuestras competencias lectoras, complementar nuestra competencia comunicativa e iluminar nuestra vida más allá de las aulas (...)” (2007: 19) y hace más concreta esta afirmación cuando señala como finalidades o razones de la enseñanza y aprendizaje de la literatura el que esta contribuye a conformar un nuevo tipo de ciudadanía que entiende un mundo global y diverso; el que la literatura llega adonde otras disciplinas no llegan y permite adentrarse y profundizar en lo más hondo de la condición humana y, por último, el que la literatura puede ayudarnos a recobrar la palabra y la imaginación en un momento en el que las nuevas formas de construcción de la mirada nos las han arrebatado.

Por su parte, Carter y Long (1991: 2) en su explicación del concepto de la competencia literaria hacen hincapié en la necesidad de hacerse con aspectos tan importantes como la identificación, el gozo y disfrute con la literatura:

Students need to be able to identify and identify with the experience, thoughts and situations which are depicted in the text. They need to be able to discover the kind of pleasure and enjoyment which comes from making the text their own, and interpreting it in relation to their own knowledge of themselves and of the world they inhabit (1991: 2; cit. en Leibbrandt, 2007).

Así pues, estaríamos hablando de una concepción de la literatura que la convierte en motor para el crecimiento personal y el cambio de la sociedad, así como para la elaboración de un juicio crítico y una palabra propia. A esto, podemos añadir, además, su valor como fuente de conocimiento, como instrumento para desarrollar la imaginación y la creatividad, y para activar el pensamiento o, como ya se venía viendo desde el XIX, como modelo de lengua. En conclusión, cuando hoy enseñamos literatura, estamos intentando que nuestros alumnos aprendan a pensar y a ver, que descubran los mensajes explícitos e implícitos que se encuentran en los textos, y puedan elaborar un juicio crítico sobre ello, así como que el reflejo de lo que leen les lleve a ahondar en su propia condición humana y a encontrar su propia palabra, sin olvidar, por supuesto, que debemos enseñarles a gozar de uno de los mejores y más placenteros reductos de paz que van a encontrar.

Precisamente para esto, siguiendo lo que ya se había iniciado en EE.UU. en los años cincuenta y con el impulso de autores como Rodari, en los años setenta, comenzó a cobrar importancia la creatividad en las clases de literatura, y se empezaron a generalizar los talleres literarios en las aulas. Según Delmiro Coto (2001), esto permitiría:

...enseñar a nuestros alumnos a ser capaces de interactuar con los textos (...) que tengan delante, a valorar la trascendencia cultural de sus significados en el momento histórico de su aparición, y el sentido –y manejo ideológico– que siguen teniendo en nuestra vida cotidiana. (...) En suma, los acercáramos a la literatura desde el lado de quienes la producían como un texto mucho más despreocupado de su pervivencia histórica que dotado de intención expresiva y de guiños dirigidos a despertar en la conciencia del receptor (2001: 2).

Es decir, se trataría de integrar en las aulas lo que se ha hecho fuera del ámbito escolar en talleres de escritura en los que, frente al teoricismo, formalismo e historicismo que, según Delmiro Coto (2001: 29), constituyen la columna vertebral de la formación universitaria de los docentes, se ha ofrecido el "placer de escribir" con "actividades de escritura, lectura e interpretación bien integradas entre sí que, sin relegar el protagonismo del autor del texto, cuenten con el papel del receptor y usuario del mismo" (29). Es decir, se ha dado la mano a las teorías de la recepción y se ha fomentado el desarrollo de una mente creativa en el alumnado, ya que, siguiendo las ideas de Rodari, creatividad es sinónimo de pensamiento divergente y una mente creativa es capaz de elaborar juicios autónomos e independientes.

Sin embargo, ya en 2001, Delmiro advierte: "Nos hallamos en una época en la que el horizonte de expectativas de los adolescentes aleja de sí cada día más al texto escrito. Su comportamiento se corresponde con el de auténticos 'depredadores audiovisuales' que consumen usos comunicativos provenientes

de la televisión, de la publicidad, del cómic, de los videojuegos, del ordenador y de Internet". Es decir, como decía al principio de este trabajo, la realidad de un mundo digitalizado se impone y, si queremos, verdaderamente, desarrollar la competencia literaria y la creatividad en nuestros estudiantes, no podemos obviarlos.

Varios artículos han abordado ya este tema. Todos ellos parten de la necesidad de renovar los métodos de enseñanza de la literatura y apuntan a las nuevas tecnologías para conseguirlo. Así, Joaquín Aguirre habla de "innovación" vinculada a la enseñanza de la literatura y señala que las nuevas tecnologías permiten trabajar con un concepto de literatura que trasciende la idea de una "adición de conocimientos" para llegar a la idea de Literatura como espacio de experimentación dinámico, como una actividad y no sólo como una acumulación de conocimientos (2002). Según este autor, el cambio de formato hace al texto maleable y plástico, es decir, manipulable, lo que, tal como recomienda Moreno Arteaga (2005) lo convierte en un "juguete" con el que trabajar y aprender. Asimismo, Aguirre (2002) señala las posibilidades del hipertexto en cuanto a la capacidad de establecer asociaciones explícitas o evidenciar todos los guiños culturales o sociales del texto por medio de hipervínculos que remitan a su explicación. A esto, podríamos añadir la propuesta Álvaro Llosa² (2013) de utilizar los hipervínculos para remitir a diferentes partes del texto al asignar fragmentos textuales a palabras del texto que hagan alusión a ellos. Llosa habla de "papiroflexia" pues se trataría de deconstruir la obra, organizar las lexías como textos y establecer recorridos múltiples. Sin embargo, desde mi punto de vista, se trata también de algo más, pues, al hacerlo estaríamos "jugando" con el análisis de texto para crear una hiperficción a partir de un texto analógico. Es decir, al asignar a una palabra del texto otro fragmento al que se accedería por un hipervínculo, estaríamos realizando un análisis de texto que daría lugar a una obra digital como las de Dora García, por ejemplo *HeartBeat*.

Por su parte, Isabella Leibrandt (2007) muestra cómo las nuevas tecnologías se han integrado en la didáctica de las instituciones alemanas, tanto como medios de enseñanza, como en su faceta de objetos de reflexión teórica, y presenta varios ejemplos, entre los que destacan los trabajos con soporte en web. Por supuesto, la creación de plataformas o páginas web sobre autores o sobre la literatura de una determinada región, como las que ella señala, permiten la innovación que exigía Aguirre y fomentan la creación del aprendizaje por parte del alumnado, así como la movilización de diferentes destrezas y aptitudes para conseguirlo; sin embargo, seis años después de este artículo, dada la vertiginosa velocidad con la que avanzan las nuevas tecnologías, son muchas más las posibilidades que tenemos a nuestro alcance, de manera que, además de crear páginas webs, disponemos de herramientas que nos permiten transformar la enseñanza de la literatura en la doble vertiente de la que hemos venido hablando: innovación y creatividad. Asimismo, las posibilidades que nos ofrecen las redes sociales, como Facebook o Twitter nos permiten entrar plenamente en la cotidianidad de nuestros estudiantes y presentarles allí mismo la literatura como algo cercano que forma parte de su vida.

² Propuesta presentada en la comunicación "Nanoliteratura: editando la ficción impresa en una nueva dimensión", presentada en el *III Congreso Internacional de BETA: Horizontes Literarios: La Obra y su entorno* en mayo de 2013.

Aquí empieza la aportación que quisiera hacer con este artículo, en el que, a partir de experiencias realizadas por el grupo Leethi (<<http://pendientedemigracion.ucm.es/info/leethi/>>) durante los últimos años, y de prácticas del aula de "Educación y Literatura" de la Facultad de Educación de la UCM, presentaré una serie de herramientas y programas que pueden ayudarnos a enseñar literatura de otra manera, es decir, a transmitir literatura (tal como lo entiende Artega) y a guiar a nuestros alumnos en su aprendizaje. Con algunas de ellas se puede trabajar, sobre todo, la innovación en la enseñanza, mientras que otras favorecen principalmente la escritura creativa con el incentivo, ya señalado por Leibbrandt, de la publicación de los textos en red y la culminación de la comunicación literaria al llegar al lector.

En cuanto a la apuesta del grupo Leethi por una enseñanza de la literatura que se adapte al siglo XXI, queda patente en las publicaciones que han visto la luz en los últimos años. Así, desde el proyecto de *Guías de lectura*³, en 2002, se ha trabajado con la necesidad de la inserción de las herramientas informáticas en las aulas. De acuerdo con esto, en 2010, se empezó a trabajar en *Alicia a través de la pantalla. Lecturas literarias en el siglo XXI* (Goicoechea y García Carcedo [eds.], 20013), obra en la que el grupo se centró en el análisis de la lectura en pantalla y estudió desde los nuevos rituales de lectura y las pautas ergonómicas más recomendables (Goicoechea, Sanz) hasta las ventajas cognitivas que ofrecen una lectura y una literatura que dejan de ser meramente visuales para apelar a otros sentidos como el tacto, o el oído (Regueiro), entre otras cosas.

Por otro lado, desde la Facultad de Educación de la Universidad Complutense de Madrid, la preocupación de Pilar García Carcedo por la integración de la escritura creativa en la formación de los nuevos lectores, a partir de las ideas de Rodari, se ha visto reflejada en la publicación de libros como *Educación literaria y escritura creativa* (2012) o *El reto de escribir* (García Carcedo y Regueiro Salgado, en prensa).

Ambas inquietudes han sido las responsables de la trayectoria que ha seguido el grupo y de las prácticas docentes que voy a presentar a continuación.

En primer lugar, como último hito del trabajo conjunto de los grupos Leethi e ILSA de la UCM, hace dos años, se comenzó a desarrollar la herramienta de anotación colaborativa @note. Con esta herramienta, se busca facilitar la anotación (y, por tanto, el análisis) de obras literarias en formato digital, de manera que estas anotaciones puedan recuperarse por categorías para facilitar la elaboración de un documento escrito, en principio de tipo académico, tal como un ensayo o una edición crítica. Dentro del aula, la herramienta sirve para enseñar o guiar a los alumnos en su viaje a través del libro. Para ello, el profesor debe crear unas categorías adecuadas a aquello que quiere resaltar del texto, de manera que los alumnos, al leer y anotar lo que ven, etiqueten su comentario (del mismo modo que lo hacen en los blogs o chats) con una de estas categorías, lo que les ayudará a conceptualizar y categorizar el conocimiento, a la vez que favorece la recuperación de las comentarios etiquetados para su posterior utilización en la redacción de un trabajo. Así,

³ <<http://pendientedemigracion.ucm.es/info/leethi/nav/texto.php?file=../docs/Una%20did%20E1ctica%20en%20hipertexto%20para%20la%20literatura,%20hacia%20la%20autonom%20dirigida.html>>.

por ejemplo, podría pedirse a los alumnos un análisis tradicional con etiquetas como temas, personajes, etc. o uno más específico, con etiquetas relacionados con una corriente crítica específica. Por otro lado, el hecho de que el libro se trabaje en modo imagen posibilita la anotación de ilustraciones lo que permite el trabajo con álbumes, libros ilustrados, cuentos infantiles o, en un futuro, obras digitales, en las que abordar la lectura de imágenes, la écfrasis, etcétera.

Por otro lado, el trabajo en las aulas nos obliga a utilizar también herramientas y programas ya disponibles y accesibles para todos los alumnos, lo cual, lejos de ser una limitación, se convierte en una oportunidad pues, en este caso, tenemos de nuestro lado la ya citada pedagogía de la rutina, que permitirá que nuestros alumnos desarrollen mejor sus capacidades dentro de un entorno que les es conocido y favorable. En efecto, existen muchos programas que pueden resultar de enorme utilidad tanto en la innovación y la apuesta por el aprendizaje activo, como en la potenciación de la escritura creativa. Algunos de ellos son ya conocidos por todos y no precisan una mayor explicación (el *power point*, por ejemplo), pero otros, aun siendo conocidos, sí requieren una mínima reflexión para ver de qué modo podemos utilizarlos en el aula, cuáles son sus ventajas y cuáles sus inconvenientes.

- *Blogs*

El blog es un buen primer ejemplo, dado que se presta tanto a la innovación como a la creatividad según el enfoque que se le quiera dar. Frente a las páginas webs que antes mencionamos, los blogs tienen la ventaja de su fácil manejo (apto para casi cualquiera) y permiten integrar multitud de recursos, tanto los que el propio *blogger* ofrece (presentaciones, imágenes, videos, etc.), como los que se pueden integrar por medio de enlaces (*bitstrips*, *moviemaker*, *YouTube* y casi cualquier programa). Otra de sus ventajas es el fácil acceso a los materiales necesarios para poder utilizarlo: tan sólo serían precisos ordenador y un proyector, recursos disponibles en casi todos los centros.

Por otro lado, el uso de los *blogs* puede ayudarnos a crear hábitos de lectura y a enriquecer el intertexto lector y, por ende, a crear lectores competentes. En primer lugar, las listas de *blogs* más leídos o "*blogs* que sigo" favorecería el desarrollo de un juicio crítico en el momento en el que el profesor prestase atención a las elecciones de los alumnos y les hiciese defenderlas y justificarlas a partir de criterios de calidad, temática, etc. Por otro lado, la posibilidad de dejar comentarios y de contestarlos fomentaría, de nuevo, el juicio crítico, así como la reflexión literaria y la utilización del lenguaje específico, es decir, estaríamos propiciando una reflexión metalingüística. Por otro lado, en el caso de los *blogs* literarios, podría suceder que la obra de uno de los alumnos inspirase a otros de modo que estos realizaran nuevos textos. Esto ayudaría a explicar las influencias literarias, la intertextualidad etc. a partir de la vivencia en primera persona. Por último, la posibilidad de elaborar una *blogonovela* con los alumnos permitiría abordar múltiples aspectos. Por un lado, podría trabajarse la reescritura colectiva de obras canónicas, como hizo Dolores Romero en el blog *La Venus Digital* (<http://venusdigital.wordpress.com/>) a partir de *La Venus mecánica* de José Díez Fernández; por otro, podrían trabajarse las características de la novela por entregas, etcétera.

En el caso de los *blogs* temáticos, la parte de investigación para llenarlo de contenido estaría íntimamente relacionada con la creación del propio conocimiento y el aprendizaje autónomo.

Por otro lado, si el *blog* se publicase en abierto con libre acceso, la posibilidad ofrecida a los padres de ver lo que están haciendo sus hijos podría favorecer la implicación familiar.

- *Glog*

Muy semejante al *blog* es el *glog*. El *glog* también es gratuito y fácil de usar y, como el *blog*, permite la inclusión de diversos recursos, como vídeos, grabaciones de sonido, etc. Lo que lo hace diferente al *glog* es el diseño, probablemente más atractivo, y la posibilidad de abrirlo como profesor y dar de alta a los alumnos.

Si nos centramos ahora en los programas o herramientas útiles para la innovación en el aula, parece innecesario ahondar en los valores educativos de la *WebQuest* para el desarrollo de un aprendizaje guiado a través de recursos procedentes de internet, puesto que ya se ha escrito bastante sobre ello.

Sí sería novedoso, sin embargo, el uso de las redes sociales con esta finalidad.

- *Facebook*

Aunque partamos de la premisa de que no todos los niños que vamos a encontrar en las aulas de secundaria pueden utilizar esta herramienta, es un hecho que los mayores, así como los alumnos de bachillerato y los universitarios utilizan plataformas como *Facebook* o *Tuenti* constantemente. Por ello, apelando una vez más a la pedagogía de la rutina, debemos convertirlos en nuestros aliados de cara al aprendizaje.

En el caso de *Facebook*, las posibilidades de uso son diversas. La más sencilla sería la creación de una página en la que los alumnos y el docente pudiesen anunciar eventos literarios (organizados por ellos o celebrados en la localidad en la que se encuentren) y compartir publicaciones en *blogs*, enlaces de páginas de autores, etc. de modo que los estudiantes recibiesen información actualizada sobre el mundo literario de hoy o sobre materiales literarios en la red. Igual que con el *blog*, en este caso, el admitir a los padres como "amigos" podría favorecer su implicación en el aprendizaje de sus hijos, así como en la organización de jornadas, etcétera.

Por otro lado, desde un enfoque más creativo, Facebook nos permite jugar y entrar en el espacio ficcional por medio del desarrollo de perfiles de personajes literarios. Así, podríamos crear perfiles de escritores, personajes destacados, etc. o, mejor aún, podríamos elegir una obra y crear el perfil de los distintos personajes, estableciendo vínculos entre ellos y haciendo que interactuasen, tal como hizo con los personajes de *Hansel y Gretel* la alumna Lara Fernández (<<http://www.facebook.com/hansel.lenador>>). De este modo, sin apenas darse cuenta, los alumnos estarían realizando un análisis profundo de los personajes, no sólo a partir de la información explícita que diese la obra en sí, sino también de todas las inferencias y deducciones que se verían obligados a hacer para poder crear el perfil y establecer una relación verosímil, con comentarios verosímiles, etc. con los otros personajes de la obra.

- Para algo parecido en cuanto a la inmersión en el mundo ficcional podríamos utilizar el *Twitter*.

Por un lado, como con *Facebook*, podríamos dar voz a distintos personajes de una obra para que *twitteasen* sus opiniones o sentimientos conforme a su carácter. Por otro lado, podríamos utilizar la plataforma como una especie de diario en el que los personajes diesen sus impresiones sobre lo que ocurre en la obra a tiempo real. Un ejemplo sería el trabajo realizado para clase por Cristina Candeias en el que los personajes de *Cuatro corazones sin freno y marcha atrás*, de Jardiel Poncela, escribían *tweets* sobre lo que iba sucediendo en la obra, de manera que la apropiación de los personajes y la reescritura eran absolutas (@4_corazones).

Una utilización menos creativa de la herramienta sería la de *twittear* alguna clase, conferencia, etc. para obligar a los alumnos a discernir cuáles son los mensajes y las informaciones principales.

Si nos centramos ahora en el desarrollo de la creatividad, podemos ver que muchas de las herramientas que hemos analizado pueden ayudarnos en este sentido: los *blogs*, *glogs*, e, incluso el *Twitter*. Asimismo, existen otras herramientas muy útiles para elaboración de cuentos, cómics, etc. o para la reescritura en estos formatos de obras literarias que pueden resultar de ardua lectura a determinadas edades. Entre estas herramientas creo que hay tres especialmente destacables.

- *Storybird*

Pensada inicialmente para la elaboración de cuentos en inglés, *Storybird* es una herramienta que puede resultar muy útil en cualquier aula de literatura. Lo que ofrece es una serie de dibujos, agrupados por tipologías y temáticas (animales, por ejemplo) y una línea de secuenciación. Se trata de que el usuario elija los dibujos que más le gusten y los vaya ordenando, según le plazca, en su línea temporal, acompañados del texto que quiera escribir para ellos. De este modo, son varios los ejercicios que podemos realizar, desde la creación de un cuento de manera absolutamente libre por parte de los alumnos, hasta la escritura guiada en la que el profesor proporcione las imágenes y los alumnos tengan que escribir el texto, o viceversa. Otro ejercicio que se podría proponer en niveles más bajos sería el de ofrecer varias viñetas (en las que el orden secuencial fuese claro) desordenadas, de manera que los alumnos tuviesen que colocarlas en orden de acuerdo con la causalidad, con lo que estarían aprendiendo a detectar las causas y a crear secuencias narrativas.

- *Bitstrips*

Otra herramienta que no podemos pasar por alto es *Bitstrips*, sobre todo, en el momento en el que se ofrece como una aplicación de Facebook y está siendo utilizada con profusión por jóvenes y adolescentes. Esta herramienta sirve para crear cómics cuyos personajes diseñan los usuarios. De este modo, podemos reformular cualquier obra en formato cómic o, incluso, crear avatares de los autores (igual que los usuarios de Facebook lo hacen consigo mismos) e inventar historias en las que se recree el ambiente literario, la época, etcétera.

- *Prezzi* interactivo

Una última herramienta destacable por su utilidad a la hora de realizar ejercicios creativos es el *prezzi* interactivo. Más allá de presentaciones, el *prezzi* interactivo nos podría servir para crear un taller literario en el que cada diapositiva se convirtiese en la "representante" de uno de los elementos del relato (personajes, temas, espacio...). Así, en ellas se ofrecerían varias posibilidades (varios espacios, varios protagonistas, etc.) para que el alumno eligiese y, una vez seleccionados todos los elementos, crease su propia historia. De este modo, además de incentivar la creatividad, estaríamos ayudando a fijar en el alumnado los elementos de la narración y, por lo tanto, introduciéndoles en el análisis textual.

Por supuesto, hay muchas más herramientas que podemos utilizar y con seguridad cada día aparecerán nuevas aplicaciones que podremos rescatar para la docencia. Hoy por hoy, todas las que hemos mencionado, así como *Skype in the class*, *Wordle* (perfecto para detectar isotopías en un poema, por ejemplo), etc. nos pueden ayudar a innovar y a favorecer el aprendizaje de nuestros estudiantes.

Nos quedaría, únicamente, explorar las posibilidades que se abren en niveles superiores, universitarios o de postgrado, en los que se puede trabajar con programas más avanzados para expresar ideas más complejas a partir de obras digitales que se inserten en lo que significa la literatura digital hoy, como campo de experimentación tecnológica y como campo en el que observar la formación de la memoria de un mundo hipercolonial⁴ con aspiraciones a global. Al igual que en los niveles inferiores, animando a los alumnos a realizar sus propias creaciones digitales, estaríamos cumpliendo objetivos como activar el pensamiento creativo y movilizar el intertexto lector para producir obras de cuya continuidad e inserción en la tradición literaria los mismos estudiantes son conscientes. Este acercamiento real de los alumnos de máster y doctorado a la literatura digital es lo que se está haciendo en el Erasmus intensivo *European Digital Literatures*, que se viene celebrando en la UCM desde julio de 2012 (<<http://digitalliterature.eu/>>). En él, los estudiantes trabajan y aprenden de autores y especialistas en literatura digital procedentes de diferentes países (Rui Torres, Philippe Bootz, etc.) y realizan sus propias obras digitales de manera que puedan comprender la esencia de un tipo de literatura producto de los tiempos en los que vivimos. Un ejemplo de las obras llevadas a cabo por los estudiantes sería *Flowing Scenes*, de D. Benito, M. Gila, A. Noain y V. de Loof (<<http://digitalliterature.eu/erasmus/projects/flowing/flowing-scenes.swf>>).

Tras este recorrido por diferentes niveles educativos y por diferentes propuestas en las que las nuevas tecnologías dan la mano a la atemporal literatura, podemos concluir que en un mundo digital, en el que, además, hemos descubierto que la mejor manera de enseñar es potenciar el aprendizaje y la mejor manera de aprender la literatura es sentirla y entenderla; en un mundo en el que la literatura se ha convertido en pieza clave para el desarrollo del pensamiento crítico, es fundamental que los profesores de literatura no desaprovechemos todas las oportunidades que nos brindan las nuevas tecnologías. Estas nos permitirán innovar en la transmisión de la literatura en todos los niveles, pues nos

⁴ Término acuñado por Amelia Sanz (2013).

ofrecerán desde la posibilidad de “jugar” con la literatura, hasta fomentar el aprendizaje autónomo con *webquest* o *blogs* que guíen el aprendizaje y favorecer la identificación con los personajes por medio de la inclusión de estos en sus redes sociales; hasta, en última instancia, impulsar la reflexión metaliteraria y la solvencia digital en los alumnos de tercer ciclo. En cualquier caso, se tratará siempre de fomentar la aparición de mentes creativas, en los términos de Rodari, es decir, de mentes capaces de pensar de un modo divergente y de romper los esquemas impuestos por experiencias previas.

BIBLIOGRAFÍA

- Aguirre Romero, Joaquín M^a, “La enseñanza de la Literatura y las Nuevas Tecnologías de la Información”, *Especulo*, 21 (2002), <<http://pendientedemigracion.ucm.es/info/especulo/numero21/eliterat.html>> [10/05/2013].
- Colomer Martínez, Teresa, “La didáctica de la literatura: temas y líneas de investigación e innovación”, en Lomas, Carlos (coord.), *La educación lingüística y literaria en la enseñanza secundaria*, Barcelona, Universitat de Barcelona, Instituto de Ciencias de la Educación, 1996, pp. 123-142, *Biblioteca Virtual Miguel de Cervantes*, <<http://www.cervantesvirtual.com/obra-visor/la-didactica-de-la-literatura-temas-y-lineas-de-investigacion-e-innovacion/html/>> [27/05/2013].
- Delmiro Coto, Benigno, “La escritura en los aledaños de lo literario”, *Tarbiya*, 28 (2001), pp. 9-50.
- García Carcedo, Pilar, *Educación literaria y escritura creativa*, Granada, Greu Ediciones, 2011.
- Goicoechea, María y García Carcedo, Pilar (eds.), *Alicia a través de la pantalla. Lecturas literarias en el siglo XXI*, Salamanca, Fundación Germán Sánchez Ruipérez, 2013.
- Gubern, Román, *Metamorfosis de la lectura*, Barcelona, Anagrama, 2010.
- Jover, Guadalupe, *Un mundo por leer*, Barcelona, Octaedro, 2007.
- Leibrandt, Isabella, “La didáctica de la literatura en la era de la medialización”, en *Especulo*, 36 (2007), <<http://www.ucm.es/info/especulo/numero36/didalite.html>> [15/05/2013].
- Moreno Arteaga, Jorge, “De la didáctica de la Literatura a la transmisión de la Literatura: Reflexiones para una nueva educación literaria”, *Especulo*, 31 (2005), <<http://www.ucm.es/info/especulo/numero31/didactic.html>> [11/05/2013].
- Rodari, Gianni, *Gramática de la fantasía. Introducción al arte de contar historias*, Barcelona, Planeta, 2002.
- Sanz Cabrerizo, Amelia, “Digital Humanities or Hypercolonial Studies?”, *RICT Responsible Innovation*, 2013, <<http://torrii.responsible-innovation.org.uk/page-resource/Ethical%20Issues>> [27/07/2013].



RESUMEN

Avanzada ya la segunda década del siglo XXI, parece que se hace necesario escuchar las voces que, desde hace varios años, han abogado por un cambio en la forma de enseñar literatura. Teorías como la constructivista nos indican los mejores sistemas para propiciar el aprendizaje y propuestas como la pedagogía de la rutina de Román Gubern (2010) nos ponen en la pista de qué hacer para adaptar nuestra enseñanza a la sociedad “pantallizada” del siglo XXI, pues, como afirma, aprendemos mejor cuando lo hacemos a partir de la realidad que nos rodea y a la que ya estamos acostumbrados.

Uniendo estas teorías a la defensa de la creatividad y de la escritura creativa, iniciada por Gianni Rodari, parece que las mejores formas de enseñar literatura hoy serán las re-

creaciones literarias y las investigaciones que utilicen como herramientas las nuevas tecnologías. Así pues, a partir de estos principios y de su aplicación en el aula, en este trabajo mostraré algunas de las posibilidades que nos ofrece el mundo digital para acercar la literatura a nuestros alumnos de primaria, secundaria, bachillerato, grado y postgrado de forma innovadora, amena, participativa y creativa gracias a programas y plataformas como *Blogger*, *Facebook*, *storybird*, *bitstrips*, *wordle*, *wix*, etcétera.

Palabras clave: Didáctica, literatura, nuevas tecnologías, escritura creativa, innovación.

ABSTRACT

Being already in the second decade of the 21st century, it seems to be necessary listening the voices that have been claiming for a new way of teaching Literature for several ages. Theories as Constructivism points out the best systems to foster the learning, and proposals as the "routine's pedagogy" by Román Gubern (2010) show us what to do to fit our teaching to the "screen" society of 21st century, because, as Gubern says, we learn better when we do it from the reality that we know and to which we are used to.

Linking these theories to the defense of creativity and creative writing, initiated by Giovanni Rodari, it looks like the best way to teach Literature today is rewriting canonical works and doing literary research using the new technologies. Therefore, in this paper I will show some of the possibilities that the digital offers us to teach Literature to undergraduate, graduate, and postgraduate students in innovating, funny, participative and creative way thanks to platforms as *Blogger*, *Facebook*, *storybird*, *bitstrips*, *wordle*, *wix*, etc.

Keywords: Didactic, Literature, New Technologies, Creative Writing, Innovation.

El proyecto MORFOGEN. Diseño de una aplicación web para visualizar familias léxicas

MARÍA JOSÉ RODRÍGUEZ-ESPIÑEIRA
Universidade de Santiago de Compostela
mjose.rodriguez.espineira@usc.es

YOLANDA LÓPEZ ARCA
Universidade de Santiago de Compostela
yolanda.lopez.arca@usc.es

1. INTRODUCCIÓN

El proyecto que presentamos con el acrónimo MORFOGEN¹ pretende contribuir al estudio de la morfología histórica del español. La hipótesis que lo sustenta es que la clasificación y definición de las palabras derivadas no debe ofrecerse de forma aislada, sino en forma de redes o constelaciones. El estudio de dichas redes puede abordarse desde dos puntos de vista: el diacrónico y el sincrónico, y no tiene por qué haber coincidencia entre las interpretaciones, los análisis y los resultados obtenidos desde ambas perspectivas. El proyecto tiene utilidad desde dos puntos de vista: teórico y aplicado.

1.1. Como se ha señalado, las relaciones derivativas establecidas en un estudio descriptivo sincrónico no coinciden necesariamente con las relaciones derivativas consideradas históricamente. Hacer visibles estas relaciones genéticas es uno de los objetivos de este proyecto. En Pena (2012) puede verse un detallado análisis de la secuencia derivativa Verbo → Nombre deverbal y de los reajustes que se han producido en esta secuencia como resultado de las lagunas existentes en las series latinas y de las nuevas creaciones romances. Para ilustrarlo con un pequeño ejemplo, extraído del mismo texto, los sustantivos *canto*, *traslado* y *compra* son derivados verbales ya que indican la acción de *cantar*, *trasladar* y *comprar*, respectivamente (cf. *escuchamos el canto de los ruiseño-*

¹ La participación en el congreso HDH2013 ha sido posible gracias a una subvención de la Xunta de Galicia (*Desarrollo de información en una aplicación web para el estudio morfogenético del léxico*. Ref. 10PXIB204249PR. Dirección Xeral de I + D. Período: 10/08/2010 a 30/09/2013). El proyecto global ha contado también con financiación estatal (*La red de morfología genética en el NDHE*. Ref. FFI2008-03532. Ministerio de Ciencia e Innovación. Período 01/01/2009 al 31/12/2011; *Estudio morfogenético del léxico español*. Ref. FFI2012-38550. Ministerio de Economía y Competitividad. Período: 01/02/2013 a 31/01/2016).

res, el traslado de la corte se produjo en 1808, la compra de esta casa será un negocio). Sin embargo, al estudiar en paralelo las series derivativas latinas y españolas, se comprueba que históricamente dicha relación V → N es válida únicamente para *comprar* (documentada en 1095) y *compra* (documentada en 1102²). En cambio, el verbo *trasladar* está creado sobre el sustantivo *traslado* (heredado del latín *translātus*, que a su vez es deverbial de *transferre* y este de *ferre*). Por otra parte, el sustantivo *canto* se ha asociado en español con el verbo *cantar* (que continúa el latín *cantāre*, frecuentativo de *canere*), pese a que etimológicamente *canto* proviene de *cantus*, deverbial de *canere* (*canere*>*cantus*).

(1) **SINCRONÍA**

V	→	N
cantar		canto
trasladar		traslado
comprar		compra

DIACRONÍA

latín		español
canere, cantus,-ūs	---	canto
cantāre, cantātio	---	cantar
transferre, translātus,-ūs	---	traslado, trasladar

Para los diccionarios etimológicos resulta imprescindible la identificación de las familias léxicas (Campos, 2008). El proyecto MORFOGEN tiene como otro de sus objetivos dar soporte o cobertura al NDHE de la RAE en cuanto a su organización genética.

1.2. En el campo de la didáctica de lenguas, son muchos los especialistas que han constatado las ventajas de una enseñanza sistemática del vocabulario, para lo cual resultan muy provechosas todas las estrategias basadas en la elaboración de constelaciones o racimos de palabras, sean de tipo estrictamente semántico, sean de base morfológica. En las citas recogidas en (2-4) se reconoce la utilidad de identificar las relaciones morfosemánticas en el aprendizaje del vocabulario:

- (2) The important principle behind the idea of a word family is that once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort (Bauer & Nation, 1993: 253).
- (3) Il s'agit de faire travailler l'apprenant sur les "familles" morpho-sémantiques où un lexème produit des mots de différentes catégories grammaticales, avec différents effets de sens, par le jeu des préfixes et des suffixes. Des règles de dérivation relativement simples permettent de passer du verbe *réparer* aux dérivés *réparation* (nom d'action), *réparateur* (nom d'agent) à l'adjectif *réparable*, et à son antonyme *irréparable* ou du nom *courage* à l'adjectif *courageux* et à l'adverbe *courageusement* (Picoche, 1999).
- (4) Los alumnos experimentan entusiasmo y sorpresa cuando se enteran de que *considerar* significa etimológicamente 'consultar con los astros'; *alumno* es 'el que ha de ser alimentado'; *feliz* es 'la que da de mamar'; *entusiasmarse* es 'meterse dentro de Dios' o 'tener dentro un Dios'; y *menopausia* es 'el descanso del mes'. (Campa et al., 2008: *Introducción*).

² La datación está tomada del DCECH.

Las afirmaciones anteriores respaldan, por tanto, la tesis de que las relaciones morfológicas y las etimológicas resultan imprescindibles para comprender el significado y la evolución de las palabras. Para dar respuesta a esta necesidad, en el proyecto MORFOGEN se ha creado una aplicación web (bdme.usc.es) integrada por dos módulos: a) una base de datos de morfología del español (BDME) que, gracias a su reformulación digital, permite el trabajo colaborativo de varios usuarios simultáneos así como la realización de búsquedas de datos combinados de varios tipos (§5); b) una herramienta de visualización de familias léxicas.

En esta contribución se examinan los siguientes aspectos: en el §2 se analizan las diferentes maneras de concebir la noción de *familia léxica*; en el 3 se resumen las fases del proyecto MORFOGEN entre 2009 y 2012 y las modificaciones llevadas a cabo en ese período. El §4 se dedica a la descripción de la BDME en su configuración actual y a las búsquedas que permite. El §5 está destinado a mostrar dos formas de visualización de los datos. El §6 hace referencia a futuros trabajos y posibles desarrollos de la aplicación.

2. LA NOCIÓN DE *FAMILIA LÉXICA*

Una noción central para este proyecto es el concepto de *familia léxica de palabras* (*word family*, *famille de mots*). La etiqueta de familia hace referencia a relaciones de parentesco y el adjetivo *léxico* remite a un tipo de significado. Por este motivo, una base de datos léxica puede albergar diferentes agrupaciones de palabras. Al menos son tres las posibilidades que hemos identificado:

2.1. En las denominadas bases de datos léxicas, las palabras se agrupan por relaciones semánticas. Constituyen lo que también se conoce como "campos semánticos"³; una versión antigua de estas agrupaciones la encontramos en los diccionarios de ideas afines, y parcialmente en los de sinónimos y antónimos. Una aplicación web para el inglés que trabaja con asociaciones de significados y de conceptos es *VISUWORDS Online Graphical Dictionary*, donde las diferentes relaciones se marcan con diferentes trazos de líneas y colores, y cuya representación pretende imitar redes neuronales. Entre las relaciones que considera tienen cabida las siguientes: (i) es un tipo de; (ii) es un ejemplo de; (iii) es un miembro de; (iv) es una parte de; (v) es una sustancia de; (vi) es similar a; (vii) pertenece a; (viii) participio; (ix) atributo; (x) se opone a; (xi) grupo verbal; (xii) implica, etcétera.

Como puede deducirse fácilmente de la lista, se trata de relaciones muy heterogéneas, ya que se combinan asociaciones conceptuales (algunas próximas a la heteronimia, la sinonimia o la antonimia), con relaciones lógicas (implicación) y con conexiones de carácter gramatical (participio, grupo verbal). Por ello, los resultados obtenidos pueden ser muy útiles desde el punto de vista conceptual, pero ofrecen menos interés desde un punto de vista puramente lingüístico. Debemos admitir, no obstante, que la forma de plasmar gráficamente estas relaciones ha servido de inspiración a los informáticos que han diseñado nuestra herramienta de visualización.

³ También puede consultarse el proyecto *Wordnet*, cuya idea central es la organización del léxico en campos semánticos (<<http://wordnet.princeton.edu/>>).

2.2. La base de datos francesa POLYMOTS se apoya en un concepto morfofonológico de familia⁴: a partir de una forma radical o palabra base, que se considera una unidad recurrente y que puede tener o no significado en la lengua, se agrupan las palabras derivadas. Sobre dicha forma radical se pueden crear palabras transparentes, como las que originan *bras* 'brazo', *fil* 'hijo' o *table* 'mesa', ya que combinan realización fónica y sentido, o bien palabras opacas, como *ferv-*, *duct-*, *oval-*, que no se asocian a un contenido en francés actual. La presencia de estos elementos se debe a que las autoras de esta base de datos, expertas en procesamiento del lenguaje natural, han constatado que una agrupación limitada a relaciones formales transparentes dejaba sin atender buena parte del vocabulario francés de origen culto (palabras como *conduire*, *conducteur*). Por este motivo, han agrupado bajo un radical opaco como *duct/du* un conjunto de 70 palabras que guardan alguna relación con la latina *ducere* (*duct-* es la forma de *perfectum* latino y *-du-* es la forma que aparece en algunos derivados franceses como *conduire*). El inconveniente de este análisis es que, sin atender a la diacronía, se propone segmentar en sincronía el verbo *conduire* del siguiente modo: prefijo *con-*, radical *-du-* y desinencia *-ire*⁵.

2.3. El tercer punto de vista es el que realmente nos interesa. Se trata del concepto de *familia léxica*, que tiene en cuenta relaciones morfo-semánticas o relaciones derivativas. Ahora bien, existen al menos dos formas de concebir las relaciones morfo-semánticas.

2.3.1. Adoptando una perspectiva exclusivamente sincrónica, se considerarán parte de una misma familia léxica todas aquellas palabras que posean semejanza formal de la que deriva cierta proximidad significativa. Este enfoque tiene la ventaja de que permite abarcar términos pertenecientes a diferentes clases de palabras. Existe una aplicación web, desarrollada por el Grupo de Estructuras de Datos y Lingüística computacional de la Universidad de Las Palmas de Gran Canaria, que lleva a cabo esta tarea y representa las palabras del español en una especie de árbol, que enlaza la palabra cabeza con sus derivados (se denomina "detector automático de relaciones morfológicas derivativas en el léxico del español"). En palabras de los autores de esta aplicación (Santana *et al.*, 2013):

Como fruto del proceso formativo que ha sufrido una palabra, se le asocian sus relaciones morfológicas, se define el mecanismo por el cual se establecen y su regularidad en la formación. Se recopilan todos los procesos formativo-derivativos del español sin entrar en discusiones teóricas de uso, formalismos históricos y otros aspectos poco prácticos desde el punto de vista del procesamiento del lenguaje natural dirigido a la automatización de procesos lingüísticos, sin obviar, en casos puntuales, los procesos histórico-etimológicos de las palabras.

Hemos realizado una búsqueda de las relaciones morfológicas de la palabra *negro*. Además de la página que muestra el listado de sus relaciones derivati-

⁴ Vid. Gala y Rey (2008).

⁵ El proyecto fue diseñado para construir las familias de forma semiautomática, a partir de una lista de palabras extraídas del diccionario Larousse 2000 y del TLFi. Según declaran sus realizadoras, el objetivo inicial era fundamentalmente pedagógico: aprendizaje del vocabulario y de la ortografía del francés, pero en la actualidad se busca que tenga utilidad para la comunidad lingüística y para quienes trabajan en tratamiento automático de lenguas (Gala y Rey, 2008).

vas, también se puede obtener el grafo asociado. Esta aplicación ofrece varios resultados valiosos: indica la categoría léxica de la palabra (sustantivo, adjetivo, verbo, etc.), señala con colores los cambios de categoría en el proceso derivativo (el color rosa marca el cambio de adjetivo a verbo, el amarillo de adjetivo a sustantivo) y separa los afijos de cada palabra derivada. La familia que representa el árbol de las relaciones morfológicas de *negro* está formada por un total de 25 términos, con el siguiente reparto por categorías léxicas:

- (5) S = greno, negrada, negralla, negrería, negrerío, negrismo, negritud, negrizal, negror, negrura = 10
 A = negral, negroero, negrestino, negreta, negrilla, negrito, negrizco, negroide, negruzco, nigérrimo, seminegro = 11
 V = negrear, negrecer, negregar, denegrir, ennegrecer = 5

Los autores de esta aplicación reconocen la necesidad de ampliar las relaciones morfo-léxicas más allá de las estrictamente morfosemánticas, incluso desde la óptica de los análisis automáticos. En palabras de Santana y otros (2013): "La potencialidad del tratamiento automático de la lengua quedaría cercenada si las relaciones que se establecen entre las palabras se restringieran a las puramente morfológicas. Se considera de interés ampliar la concepción de relación morfológica –relación morfológica extendida– en lo que atañe tanto a la raíz como a los afijos". Como consecuencia de ello, proponen extrapolar el concepto de afijo a elementos que "no siendo tales, actúan como si lo fueran". Ello implica introducir la perspectiva diacrónica en el análisis, al incluir casos como los de *lácteo*, *osificar*, que mantienen con sus respectivas bases *-leche*, *hueso*– relaciones semánticas codificadas en la morfología, como las de 'perteneiente o relativo a' y 'hacer o convertir', respectivamente. El problema es que también se incluyen como afijos segmentos puramente fonológicos (en palabras descendientes del árabe, segmentan el artículo *al-* o la terminación *-alle* en vocablos procedentes del francés o del catalán).

2.3.2. Adoptando una perspectiva diacrónica, la familia se contempla desde una concepción genética o etimológica y se concibe como un "conjunto de palabras agrupadas en torno a una raíz, étimo o palabra primitiva, como punto de arranque de las demás palabras emparentadas" (Pena y Campos Souto, 2009: 32).

La BDME responde a esta concepción, ya que pretende servir como estructura-soporte de un diccionario histórico. Desde una perspectiva histórica, la familia léxica es una agrupación de palabras más amplia o abarcadora que otra concebida desde el punto de vista sincrónico. Tiene que agrupar todas aquellas palabras que directa o indirectamente remiten a una misma raíz, sean estas variantes cultas, populares o préstamos e independientemente del mayor o menor grado de transparencia formal y semántica que plasmen en cuanto a la relación derivativa.

Como consecuencia de la aplicación de este enfoque, se pueden obtener series heterogéneas de derivación, donde se muestran las alternancias forma popular-forma culta (cf. la macrofamilia de *hacer*, que consta de, al menos, 650 voces, con la que ejemplifican Pena y Campos, 1999: §3). Los grados de parentesco son variados –algunos más evidentes o inmediatos y otros menos evidentes–, pero siempre se pueden establecer capas o estratos que permitan identificar subfamilias dentro de esas familias más amplias hasta llegar a series de derivación más homogéneas formal y semánticamente.

3. FASES DEL PROYECTO

La primera fase del proyecto consistió en la normalización y adaptación de una base de datos previa en formato ACCESS (con información sobre latín, español y préstamos). Esta base acumulaba toda la información de las diferentes lenguas en una tabla única y codificaba con letras los distintos valores asignados en cada uno de los campos a cada nuevo vocablo analizado. La figura 1 deja ver algunos campos de la base y la codificación alfabética de las propiedades de las palabras en cada campo.

PALABRA	BASE	CLJ	CLA	PROCES	LENGUA	PALABRAE	BASEE	CLASEP	CLASE	PRE	CLB	PRÉSTAMO	LENGUAE	PRÉSTAMOE	LENGUAE	HIPOTETI
uermis,s		s		000		verme		s		000	00					
ossumi,s		s		000		hueso		s		000	00					
plumbum,l		s		000		plomo		s		000	00					
stoppa,ae		s		000	g	estopa		s		000	00					
plaga,ae	plangere	s		000		llaga		s		000	00					
orphanus,l		s		000	g	húrfano,a		n			00					
torqueze		v		000		torcer		v		000	00					
plorare		v		000		llorar		v		000	00					
pratum,l		s		000		prado		s		000	00					
uentus,l		s		000		viento		s		000	00					
pellis,s		s		000		piel		s		000	00					
latrare		v		000		ladrar		v		000	00					
niger,gramm		a		000		negro,a		a		000	00					
latro,onis		s		000		ladrón,a		n		000	00					
pluere		v		000		llover		v		000	00					
nasci:		v		000		nacer		v		000	00					
torris,s		s		000		torre		s		000	00					
rodere		v		000		roer		v		000	00					
pectus,oris		s		000		pecho,l		s		000	00					
querere		v		000		querer		v		000	00					
plangere		v		000		plañir		v		000	00					
nomen,inis		s		000		nombre		s		000	00					
orbis,s		s		000		orbe		s		000	00					
mundus,a,um		a		000		mondo,a		a		000	00					
saccus,l		s		000	g	saco		s		000	00					
telumus,a,um		a		000		ayuno,a		a		000	00					
littera,ae		s		000		letra		s		000	00					
ordo,inis		s		000		orden		s		000	00					
taurus,l		s		000		toro		s		000	00					
mens,l,s		s		000		mes		s		000	00					
laqueus,l		s		000		lazo		s		000	00					
tussis,l,s		s		000		tos		s		000	00					
novem		d		000		nueve		d		000	00					
miscere		v		000		mecer		v		000	00					
nebula,ae		s		000		niebla		s		000	00					

Fig. 1. La BDME con su formato en Access.

En esta base los campos 2 a 9 se referían a palabras latinas. Los demás a palabras del español actual. Estaba construida sobre una matriz de rasgos que proporcionaba información completa sobre los aspectos formales, funcionales y semánticos de la palabra: (i) Clase y subclase de la palabra base y de la derivada; (ii) Tipo formal de la palabra base y de la derivada: popular/culto; (iii) Tema de palabra base inexistente como palabra; (iv) Estructura mórfica de la palabra; (v) Procedimientos morfológicos utilizados en la creación de la palabra derivada; (vi) Serie de derivación de la que forman parte la palabra base y la palabra derivada; (vii) Palabra de creación autóctona, frente a palabra incorporada como préstamo. Distinción entre préstamo inmediato y mediato; (viii) Marcas de uso. Cada campo llevaba asociado un conjunto de marcas (símbolos en forma de letras) que permitían llevar a cabo la clasificación de cada una de las palabras de la base.

La remodelación de esta base de datos con vistas a su conversión digital se realizó en dos etapas. En la primera, se configuraron 12 tablas, las principales para el español y el latín, más otra para griego y una específica para temas greco-latinos (tipo *arboricida*, *adenopatía*, etc.), así como tantas tablas como lenguas intervienen, de forma recurrente, en la incorporación de préstamos. Cada tabla contenía el mismo tipo de información para todas las lenguas consideradas, a saber portugués, gallego, occitano, catalán, francés, italiano, inglés. Existía, por último, una tabla específica para otras lenguas, con menos propiedades. Al desarrollar la aplicación se constataron dos problemas: a) La tabla para otras lenguas se había simplificado en exceso y no permitía el análisis de préstamos de lenguas como el árabe o el alemán (términos provenientes del árabe antiguo o del alto alemán, por ejemplo); b) La tabla para temas greco-latinos, no se había vinculado a lenguas específicas, por tratarse de temas de palabras inexistentes. En el año 2010 se introdujeron algunos temas a partir del Diccionario Vox, con vistas a una posterior asignación como bases de palabras españolas. Al comprobar que el vocabulario científico se incorpora a través del francés y del inglés, se vio la necesidad de vincular muchos temas a estas dos lenguas; además, en las propias lenguas griega y latina hay temas de palabras inexistentes como palabras simples que sólo aparecen en cuanto constituyentes de palabras afijadas y compuestas.

Estos problemas obligaron a reconsiderar el proceso de almacenamiento de datos y sus relaciones. Desde marzo de 2012 existe una tabla única, en la que las lenguas se seleccionan mediante un menú desplegable, con el siguiente orden: a) las más usadas en el proyecto; b) las restantes por orden alfabético. El formulario de introducción de datos también se modificó, para dar cabida a campos como SIGNIFICADO, TIPOS DE AFIJOS, EJEMPLOS, DATACIÓN, etc. Al analizar una palabra, el investigador debe seleccionar entre clasificarla como subclase de palabra (adjetivo, verbo, etc.) o como tema.

Se ha dedicado especial atención en el proyecto a los siguientes aspectos: a) relación de lenguas posibles como fuentes de préstamo para el español; b) relación de sufijos y prefijos, establecida siguiendo pautas diacrónicas; c) significados morfológicos (una lista cerrada de posibles significados esquemáticos de las palabras derivadas); d) marcas de uso, que se han adaptado a las diferentes lenguas. En el caso del español, se han trasladado algunas marcas del DCECH y también se ha buscado que las marcas geográficas armonicen con las del NDHE. Se distinguen dos marcas de tipo histórico, una que sigue las convenciones del DRAE (antigua, desusada, poco usada) y otra que sigue las agrupaciones del NDHE.

4. CONFIGURACIÓN ACTUAL DE LA BDME

La BDME es una base de datos relacional, es decir, los datos se almacenan en varias tablas conectadas entre sí para permitir la interconexión de los datos recogidos en ellas. En concreto, la BDME está compuesta por un total de trece tablas que se organizan en torno a una principal, la tabla TÉRMINO. Esta tabla contiene veinte campos que permiten recoger distinta información relativa a cada una de las palabras que se almacenan en la base de datos: el idioma, el lema, la clase de palabra, los sinónimos, si tiene estatus de palabra o se considera un tema de palabra inexistente, si se trata de una palabra hipotética,

las fechas de su primera y última documentación o ejemplos de su uso, entre otras propiedades.

Algunos de estos campos están relacionados con información albergada en otras tablas. Es el caso, por ejemplo, del idioma. A cada una de las lenguas utilizadas en la base de datos está asociado un color con el que se representarán en el grafo y la visualización lineal; también se indica si se trata de una lengua que aparece con frecuencia en la base de datos o no, un parámetro que se utiliza para ordenar las lenguas en el menú desplegable del formulario web empleado para introducir los datos. Es decir, cuando se introduce el idioma de una palabra en la tabla *TÉRMINO*, automáticamente la palabra en cuestión pasa a estar relacionada también con la información almacenada acerca de esa lengua en la tabla *IDIOMA*.

La información relativa a los significados morfológicos, las marcas de uso y las bases de derivación se almacenan también en tablas independientes, aunque vinculadas a *TÉRMINO* de un modo similar al de la tabla *IDIOMA*. Por ejemplo, cuando una palabra actúa como base de derivación de otra, pasa a estar almacenada también en la tabla *BASE*, junto con el proceso mediante el cual se ha formado la palabra. En aquellos casos en los que el proceso de formación de la palabra es la prefijación, la sufijación o ambas, entran en juego otras dos tablas: *PREFIJO* y *SUFIJO*, que contienen información acerca de cada afijo, como las variantes o algunos ejemplos de palabras en las que aparecen.

También las tablas que contienen las marcas de uso y los significados morfológicos tienen asociadas a su vez otras tablas, que permiten jerarquizar la información. Las marcas de uso, por ejemplo, se organizan en cuatro niveles para el español y cada tipo de marca admite subdivisiones. Esto permite dar cuenta de la inclusión de, por ejemplo, Bolivia, Ecuador y Perú como países del área andina, variedad del español de América que es una marca de tipo geográfico.

Por otra parte, se han diseñado dos buscadores de términos. El primero, básico, permite búsquedas alfabéticas por inicio, medio o fin de palabra. Este buscador fue muy importante en la primera fase de remodelación de la base, ya que se hizo una importación de datos desde la BDME en formato *ACCESS* que, por diferentes motivos, no fue todo lo exitosa que debiera. Con buscadores de este tipo fue posible saber qué palabras no se habían introducido en la importación y cuáles sí.

El segundo buscador, denominado avanzado, permite búsquedas específicas combinando propiedades diferentes y lenguas diferentes. En la figura 2 se muestra la búsqueda de palabras del *IDIOMA* español, de la categoría *SUSTANTIVO*, que tengan origen en palabras latinas [*ORIGEN*, *IDIOMA*, Latín] cuya base sea un Verbo.

El resultado de la búsqueda se presenta en la figura 3. La pantalla informa del número de resultados (casi 1.900 términos), que se muestran ordenados alfabéticamente y con las propiedades de la tabla que el usuario ha seleccionado previamente: *TÉRMINO*, *ENTRADA*, *ORIGEN* Y *BASE*. Desde esta pantalla (o las sucesivas) se puede acceder tanto a la ficha de cada término como al grafo de la familia léxica que le corresponde.

USC MORFOGEN PLATAFORMA WEB PARA EL ESTUDIO MORFOGENÉTICO DEL LÉXICO

Inicio Añadir términos **Buscar términos** Mis datos Editar web pública

Yolanda López Arca [Cerrar sesión](#)

Buscador básico - **Buscador avanzado** - Asignar prefijos - Asignar sufijos

Término

Filtro

Posición: Inicio Medio Final

Hipotética Culta Popular

Idioma: Español, Latin, Catalán, Francés, Gallego, Griego, Inglés

Clases de términos:

- Palabra Tema Indiferente
- Sustantivo Lugar Persona
- Adjetivo
- Participio
- Verbo
- Adverbio
- Preposición
- Conjunción
- Frase/Oración
- Determinativo/Pronombre

Origen Sin origen

Filtro

Posición: Inicio Medio Final

Hipotética Culta Popular

Idioma: Español, Latin, Catalán, Francés, Gallego, Griego, Inglés

Clases de términos:

- Palabra Tema Indiferente
- sustantivo
- Adjetivo
- Participio
- Verbo
- Adverbio
- Preposición
- Conjunción
- Frase/Oración
- Determinativo/Pronombre

Núm. bases: 1

Base 1

Filtro

Posición: Inicio Medio Final

Hipotética Culta Popular

Clases de términos:

- Palabra Tema Indiferente
- Sustantivo
- Adjetivo
- Participio
- Verbo
- Adverbio
- Preposición
- Conjunción
- Frase/Oración
- Determinativo/Pronombre

Procesos:

- Reduplicación Calco Alteración
- Prefijación Sufijación
- Blending Acortamiento Siglación Cruce
- Conversión flexión
- Sustitución Tema
- Diminutivo Aumentativo Despectivo

¿Qué quieres ver? ¿En qué orden?

Propiedades: Término, Idioma, Entrada, Variantes, Sinonimia, Ejemplos, Hipotética

En el resultado: 1. Término, 2. Entrada, 3. Origen, 4. Base 1

¿Cómo lo quieres ver?

Propiedades: Término, Idioma, Entrada, Variantes, Sinonimia, Ejemplos, Hipotética

En el resultado: 1. Término, 2. Entrada, 3. Origen, 4. Base 1

¿Cómo lo quieres ver?

En pantalla - Num. resultados/pág. 20

[Realizar consulta](#)

USC

Fig. 2. Buscador avanzado de términos en la BDME.

Ficha	Grafo	Término	Entrada	Origen	Base 1
		abdicación	abdicación	abdicatio	abdicar
		aberración	aberración	aberratio	aberrar
		abertura	abertura	apertūra	abrir
		abjuración	abjuración	abjuratio	abjurar
		ablución	ablución	ablutio	abluir
		abnegación	abnegación	abnegatio	abnegar
		aboliación	aboliación	abolitio	abolir
		abominable	abominable	abominabilis	abominar
		abominación	abominación	abominatio	abominar
		aborto	aborto	abortus	abortar
		abreviación	abreviación	abbreuiatio	abreviar
		abreviatura	abreviatura	abbreuiatura	abreviar
		abrogación	abrogación	abrogatio	abrogar
		absolución	absolución	absolutio	absolver
		absorción	absorción	absorptio	absorber
		abstracción	abstracción	abstractio	abstraer
		accesión	accesión	accessio	acceder
		acceso	acceso	accessus	acceder
		acechador	acechador,a	assectator	acechar
		aceleración	aceleración	acceleratio	acelerar

Fig. 3. Pantalla de resultados del buscador avanzado.

5. VISUALIZACIÓN DE LOS DATOS

5.1. En forma dinámica

Se ha diseñado una herramienta de visualización que permite la búsqueda de una familia léxica a partir de un término almacenado en la base. El programa realiza una consulta a la base de datos y representa la familia léxica mediante un grafo o árbol con sus nudos, teniendo en cuenta dos tipos de relaciones: a) la que vincula un término con su base, es decir, aquella palabra de la misma lengua de la que deriva; b) la que vincula un término con su origen, es decir, aquella palabra de otra lengua de la que es continuación, incorporación o adaptación. La herramienta distingue las lenguas con colores y tiene en cuenta las clases de palabras. En la figura 4 aparece la pantalla de la familia léxica del adjetivo *negro*⁶, en su forma abreviada, agrupando los derivados según su categoría (S = sustantivo, A = Adjetivo, V = Verbo, O = Otras). Esta agrupación en clases aparece cuando el número de derivados es superior a 8 palabras.

Los términos ocultos pueden desplegarse haciendo clic con el ratón sobre cada uno de estos círculos. Es posible volver a ocultarlos mediante un segundo clic. También pueden mostrarse u ocultarse la base o el origen de un término haciendo clic con el botón derecho sobre él. Los nudos correspondientes a cada

⁶ En la BDME se distingue entre el adjetivo *negro,-a* y su recategorización como sustantivo.

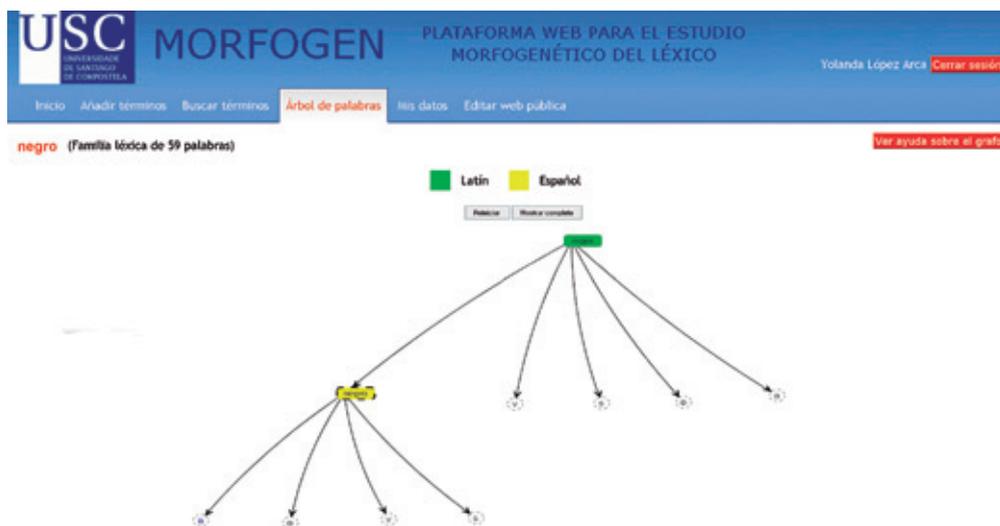


Fig. 4. Pantalla de la familia léxica de negro presentada de forma abreviada.

término pueden moverse para facilitar la visualización. Haciendo clic sobre ellos mientras se mantiene pulsada la tecla Ctrl se accede a la ficha completa en la base de datos.

En la figura 5 se ofrece la familia al completo, tanto en sentido descendente (palabras creadas en español sobre el adjetivo) como en sentido ascendente (palabras latinas originarias de las españolas). Un aspecto que necesita ser mejorado es la presentación de esta segunda imagen: los árboles de familias numerosas se muestran con ramas amontonadas y desordenadas, si bien el usuario puede mover las ramas y reubicarlas para que se visualicen mejor.

Por tanto, la representación en forma de grafo permite explorar la familia léxica de manera dinámica, profundizando en una determinada serie de derivación, eligiendo los elementos que interesen o prescindiendo de la información que no resulte interesante en una determinada consulta. Los elementos de cada lengua se representan con un determinado color que permite distinguirlos de los términos de lenguas distintas que estén implicados en la formación de la familia léxica. Las flechas que unen los términos señalan las relaciones que existen (i) entre las bases de derivación y sus derivados, (ii) entre el origen de una determinada palabra y su resultado y (iii) entre las palabras que sirven de préstamos y sus correspondientes resultados en otras lenguas.

5.2. En forma estática

La visualización lineal permite explorar una familia léxica de manera estática (Fig. 6). Para una lengua como el español reproduce de forma transparente la secuenciación de las relaciones derivativas (como $N > V > A$). Esta representación resulta especialmente útil para comprobar de manera gráfica qué palabras se han perdido en la evolución del latín al español y cuáles son creaciones hispánicas. La representación parte del término que da origen a la familia léxica, que carece tanto de base de derivación como de origen. En cada línea se muestran una a

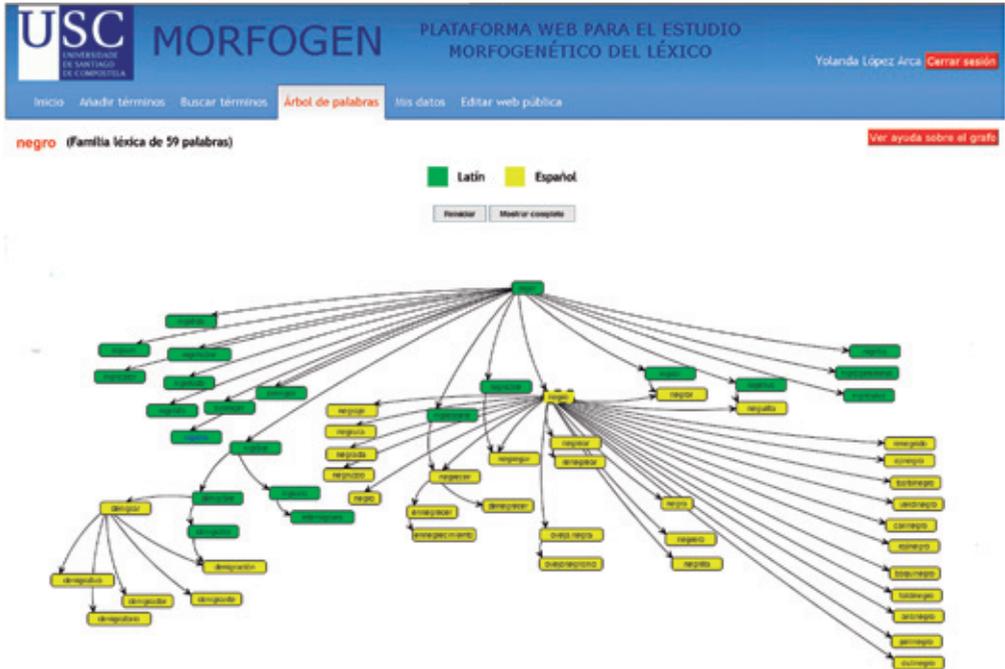


Fig. 5. Pantalla de la familia léxica del adjetivo negro, en relación con el latín niger y sus derivados. Los resultados han sido ordenados por el usuario.

una las series de derivación que forman la familia. De izquierda a derecha se representan las relaciones entre las bases de derivación y sus derivados.

En la representación pueden alternarse series de derivación en más de una lengua. Los elementos de cada lengua se identifican con un mismo color, que coincide con el del grafo. Si un término es el antecedente de otro término en una lengua distinta, su correlato aparece en la línea inferior. Las relaciones por la vía de origen se representan, pues, en el eje vertical, de arriba hacia abajo (cf. *niger-negro*, *nigror-negror*, *nigellus-neguilla*, *nigrescere-negrecer*, etcétera).

Las series de derivación en la lengua que da lugar a la familia léxica se ordenan según la clase de palabras a la que pertenezca el primer derivado de la serie. En primer lugar aparecen los sustantivos (*negror*, *negrilla*), seguidos de los adjetivos y los verbos (*negrecer*, *denegrecer*, *ennegrecer*...). Por último aparecen las palabras analizadas como adjetivos y sustantivos, adjetivos y participios, etc. En los casos en los que hay más de un derivado de la misma clase se ordenan alfabéticamente. En último lugar, en la parte inferior de la representación, aparecen las series de derivación cuyo primer término derivado se crea en español sin antecedente latino. Los criterios de ordenación se mantienen: primero se clasifican según la clase de palabras, como se indicó anteriormente, y en segundo lugar por orden alfabético (*negrada*, *negraje*, *negrura*...). Las palabras compuestas se incluyen también en este último grupo (*carinegro*, *faldinegro*...).

La representación en paralelo de las series de derivación latina y española permite observar fácilmente qué palabras se han perdido en la evolución del latín al español, ya que los huecos que ocuparían esas palabras se corresponden con

blancos en las series derivativas. Al mismo tiempo, resulta fácil explorar cuáles son las creaciones hispánicas que carecen de antecedente latino buscando las casillas de color amarillo que ocupan un lugar más a la derecha que la última casilla verde de la fila superior.

6. PERSPECTIVAS DE FUTURO

En los próximos años el equipo del proyecto se propone revisar los datos actualmente almacenados para completar el análisis con las propiedades añadidas (significados morfológicos, sufijos y prefijos, marcas de uso, datación). El vocabulario culto, construido sobre temas grecolatinos, está pendiente de análisis, debido a la remodelación del formulario. La parte etimológica de la base, propiedad ORIGEN, es susceptible de revisión y están sin completar los préstamos de otras lenguas actuales y las palabras formadas sobre temas de palabras inexistentes como palabras simples, empezando por el griego y el latín, y continuando, por este orden, con el francés, inglés y español. Se pretende, además, aumentar el número de palabras analizadas, completando en lo posible las familias léxicas.

El segundo componente de la aplicación web, la herramienta de visualización, también debe ser mejorada en las siguientes facetas: (a) evitar la acumulación de información cuando las familias léxicas son extensas, para lo cual se ha estado ensayando con el plegado y desplegado de las ramas, con el movimiento de los grafos y con la agrupación de los nudos por categorías léxicas; (b) mejorar la información asociada a cada término en los nudos de los grafos; creemos que resulta especialmente relevante añadir pequeñas glosas con el significado de las palabras y, pensando en su explotación para la enseñanza del español, añadir los equivalentes en inglés de cada término; (c) utilizar colores en las flechas para distinguir relaciones derivativas, especialmente necesarios cuando el grafo se despliega en su totalidad; (d) asociar los grafos a búsquedas específicas que permitan filtrar la información que resulte de interés para el usuario.

Una vez que la plataforma esté accesible, en función del interés que suscite, tiene una configuración que permitiría reproducir el trabajo llevado a cabo con el español para otras lenguas románicas.

BIBLIOGRAFÍA

- Bauer, Laurie y Nation, Paul, "Word families", *International Journal of Lexicography*, 6/4 (2003), pp. 253-279.
- Campa, Hermenegildo de et al., *Diccionario español escolar de familias etimológicas*, Granada, Edición experimental, 2008, <<http://es.scribd.com/doc/106187096/Diccionario-espanol-escolar-de-familias-etimologicas>> [01/07/2013].
- Campos, Mar, "Hacia la ordenación morfológica del NDHE: primer esbozo", *Verba*, 34 (2007), pp. 125-155.
- DCECH, Corominas, Joan y Pascual, José Antonio, *Diccionario crítico etimológico castellano e hispánico*, Madrid, Gredos, 1980-1991.
- Gala, Nuria y Rey, Véronique, "Polymots: une base de données de constructions dérivationnelles en français à partir de radicaux phonologiques", *Traitement Automatique des Langues Naturelles, TALN 2008*, Avignon, 2008, <http://www.florianboudin.org/taln_archives/TALN/TALN-2008/actes/taln-2008-court-010.pdf> [01/07/2013].
- NDHE, Fundación Instituto de Investigación Rafael Lapesa para el Nuevo diccionario histórico del español, <<http://www.frl.es/Paginas/default.aspx>>.

- Pena, Jesús, "Alteraciones de la serie derivativa verbo-nombre deverbal en español. Análisis genético", *Zeitschrift für romanische Philologie*, 128/2 (2012), pp. 319-349.
- Pena, Jesús y Campos, Mar, "Propuesta metodológica para el establecimiento de familias léxicas: el caso de *hacer*", *Cuadernos del Instituto Historia de la lengua*, 2 (2009), pp. 21-51.
- Picoche, Jacqueline, "Dialogue autour de l'enseignement du vocabulaire", *Études de Linguistique Appliquée*, 116 (1999), pp. 421-434, <<http://jpicochelinguistique.free.fr/ENSEIGNEMENT/articlesdefond.html>>.
- POLYMOTS, <<http://polymots.lif.univ-mrs.fr/v2/>> [01/07/2013].
- Relaciones morfológicas*, O. Santana et al., <<http://www.gedlc.ulpgc.es/investigacion/scogeme02/relmorfo.htm>> [01/07/2013].
- VISUWORDS *Online Graphical Dictionary*, <<http://www.visuwords.com/>> [01/07/2013].
- WORDNET, <<http://wordnet.princeton.edu>> [01/07/2013].



RESUMEN

El objetivo de esta contribución es presentar las características de una aplicación web destinada a hacer visibles en español las relaciones de parentesco entre las palabras de una misma familia léxica (<bdme.usc.es>). La aplicación contiene dos módulos: a) una base de datos de morfología del español y b) una herramienta de visualización. En este trabajo se mostrarán diferentes concepciones de la noción de familia léxica y también algunos procedimientos empleados en otros proyectos para dar cuenta de relaciones morfológicas. Se mostrará la conveniencia de aplicar un enfoque histórico a las familias léxicas y de trabajar con dos tipos de relaciones de parentesco: (i) la que asocia una palabra con su origen y (ii) la que vincula una palabra con su base léxica. Se comentarán algunos problemas surgidos durante el proceso de conversión de una base de datos monousuario en otra de acceso compartido. Se explicarán las principales características técnicas de la base de datos y se mostrará el diseño de dos formas de representación: en diagramas arbóreos y en forma lineal. Con la primera se pretende ofrecer una imagen global de cada familia léxica y con la segunda se busca reflejar la progresión de los procesos derivativos en las series de una misma lengua.

Palabras clave: Familia léxica, morfología histórica, etimología.

ABSTRACT

This paper shows the characteristics of a web application designed to represent the family relationships between components of a word family in Spanish (<bdme.usc.es>). The application consists of two modules: a) a morphological database of Spanish and b) a visualization tool. Different formulations of the concept of 'word family' will be shown, as well as some procedures that have been used in other projects in order to account for morpho-lexical relationships. The convenience of a historical approach to word families will be discussed, as well as the appropriateness of working with two types of family relationships: (i) the one associating a word with its origin and (ii) the one associating the word with its lexical base. Some issues that emerged during the process of conversion from a single-user to a multiple-access database will be discussed. The main features of the database will be explained. The design of two representational forms (tree and linear diagrams) will be shown: the former aims to produce a global representation of a word family while the latter seeks to represent the progression of derivational processes in the series of a given language.

Keywords: Word family, historical morphology, etymology.

Hacia la *Smartlibrary: Mnemosine*, una biblioteca digital de textos literarios raros y olvidados de la Edad de Plata (1868-1936)¹. Fase I

DOLORES ROMERO LÓPEZ
Universidad Complutense
dromero@filol.ucm.es

DELIMITACIÓN CRONOLÓGICA: DE 1868 A 1936

MNEMOSINE Biblioteca Digital de Textos Literarios Raros y Olvidados (1868-1939) abarca un período comprendido entre dos fechas cruciales de Modernidad histórica y literaria. La fecha de 1868 constituye el punto de partida ya que tras el Sexenio Democrático se desarrolla en España la novela realista, la poesía posromántica, modernista y simbolista y el teatro burgués por horas, la zarzuela y el teatro musical. La fecha de 1936 cercena las tendencias literarias modernas debido a la muerte, silencio o exilio de numerosos escritores².

PUNTUALIZACIONES SOBRE LA CATEGORÍA DE 'RAROS Y OLVIDADOS'

La primera vez que se utiliza el rótulo 'raros' como término crítico, aplicado a la literatura española es en 1863, cuando Bartolomé José Gallardo publica

¹ Esta investigación pudo iniciarse en el verano de 2012 gracias a una beca para la movilidad del profesorado universitario de la Comunidad de Madrid financiada por la Fundación Caja Madrid. Los resultados de esta investigación han recibido el apoyo del Ministerio de Economía y Competitividad. Proyecto de Investigación: "Escritorios Electrónicos para las Literaturas-2" (Ref. FFI2012-34666).

² En una versión más dilatada sobre la justificación de esta biblioteca explicaré con detalle las consecuencias de la Revolución de 1868 y del Sexenio Democrático o Revolucionario (1868-1874) para la literatura. La historia y la cultura españolas de este período es fundamental para entender la categoría de 'raros y olvidados'. Existe una lucha dialéctica e ideológica entre democracia, socialismo, anarquismo, positivismo, krausismo que representarán los valores de la llamada izquierda y la derecha que se apoya en el caciquismo, el carlismo y la iglesia. El reinado de Alfonso XIII (1902-1931) presenta un fuerte crecimiento demográfico, el desarrollo de la industria y del sector servicios y la alfabetización de hasta el 50% de la sociedad hacia 1930. El enfrentamiento entre republicanos y nacionales dividió en dos el país y tuvo consecuencias nefastas en todos los ámbitos.

su *Ensayo de una biblioteca española de libros raros y curiosos*. Es la obra de un bibliófilo, publicada póstumamente gracias a la colaboración de Zarco del Valle, Sancho Rayón y Menéndez Pelayo³. Un siglo más tarde, en 1964, Homero Serís publica *Nuevo ensayo de una Biblioteca Española de libros raros y curiosos*. Homero Serís, Vicepresidente de The Hispanic Society of America da cuenta de los fondos que se encuentran en dicha institución y que pertenecen a la colección de Mr. Archer Huntington y Mr. Ticknor⁴. Esos libros constituyen dos pilares fundamentales que muestran la preocupación desde el racionalismo ilustrado por recabar para la historia de la literatura todos los escritos, no sólo los más 'gloriosos'. Sin embargo en *Mnemosine* no se encontrarán digitalizados volúmenes anteriores a 1868. Si es cierto que el doblete "raros y curiosos" ha hecho cierta fortuna como se puede comprobar en el despliegue de referencias que ofrece el banco de datos de la Biblioteca Nacional para dichos términos, lo que interesa ahora no es tanto aquellos documentos que resultan "curiosos" sino otros tantos que se han "olvidado".

Así el concepto 'raros y olvidados' tiene en la Modernidad un significado cultural complejo referido no tanto a documentos de poca circulación o impacto, sino a aquellos que tienen una estética e ideología específicas. No se alude ya a 'libros raros y curiosos' sino a libros 'raros y olvidados'. La aparición del concepto 'olvidados' como segundo término de la categorización se lo debemos a Federico Carlos Sainz de Robles que defiende el marbete 'raros y olvidados' para hablar de la literatura española del primer tercio del siglo XX. Primeramente publica su libro *La novela corta española: Promoción de El Cuento Semanal (1901-1920)* (1959) en el que se recogen algunas de las novelas olvidadas más significativas. En 1971 vuelve sobre el tema desde una perspectiva más historicista con *Raros y olvidados: (la promoción de El Cuento Semanal*. Desde entonces la investigación sobre la promoción de *El Cuento Semanal* no ha dejado de dar frutos interesantes. Él mismo publica en 1975 *La promoción de El Cuento Semanal 1907-1925: (un interesante e imprescindible capítulo de la historia de la novela española)* y con posterioridad aparece la edición de Brigitte Magnien, *Ideología y texto en El Cuento Semanal: 1907-1912* (1986) y el más reciente de María Lourdes Íñiguez Barrera *El Cuento Semanal, 1907-1912: análisis y estudio de una colección de novelas cortas* (2005). La idea de Sainz de Robles de llamar la atención de la crítica sobre la colección de *El Cuento Semanal* a través de la apostilla de 'raros y olvidados' ha surtido, pues, el efecto esperado.

A partir de la refundación terminológica de Federico Carlos Sainz de Robles aparecen otros estudios que abundan en la necesidad de releer a autores consideramos 'menores' de este período. La categoría se convierte una etiqueta, en un cajón de sastre, en el que ir acumulando lo que el canon oficial ha ido dejando

³ Se trata de una compilación de textos de la Edad Media hasta el siglo XVII. Incluye la descripción de las obras y, en algunos casos, textos enteros o fragmentos. El orden es alfabético según el apellido del autor. En estos volúmenes se encuentran numerosos textos, colecciones anónimas de romances y fragmentos de libros de caballerías. También reproduce fragmentos de textos olvidados de autores famosos como Cervantes.

⁴ Se trata de 15.000 volúmenes en español de los siglos XV, XVI y XVII, un legado superior al que se puede encontrar en la National Library de Londres y la Biblioteca Nacional de París. Cuenta con ejemplares únicos de distintas ediciones de *D. Quijote* y un ejemplar completo de la edición de *Tirant lo Blanch* (Barcelona, 1497).

fuera. En los últimos lustros⁵ han sido bastantes las investigaciones que se han publicado sobre géneros olvidados –como por ejemplo la novela teatral (Pérez Bowie, 1996), la novela de una hora (Galindo Alonso, 2002), etc.⁶– o autores considerados de segunda fila pero que a lo largo de los años la crítica se ha ido acordando de ellos –como por ejemplo Eduardo Zamacois (Cordero Gómez, 2008), Alejandro Sawa (Correa Ramón, 2008 y Diego Fernández: 2009), Emilio Carrere (Riera Guignet, 2011) y Carmen de Burgos (Núñez Rey, 2005). También han realizado imprescindibles aportaciones sobre raros y olvidados Jaume Pont (2000), Vicente Llorens (2003), Richard Cardwell (2006), María Josep Balsach (2006), Antonio Cruz Casado (2006) y Cecilio Alonso (2008). Estos últimos utilizan la categoría ‘raros y olvidados’ para reclamar la necesaria revisión del canon de la modernidad y proponer así relecturas más abiertas, con nuevos temas, nuevos géneros y nuevas relaciones literarias nacionales e internacionales (Mainer: 1975 y Romero: 1998). Se ha venido, pues, utilizando la categoría de ‘raros y olvidados’ cuando se ha querido reivindicar la importancia histórica de algún autor o género olvidado de la literatura durante este período.

La etiqueta de “raros y olvidados” en la literatura contemporánea española tiene valor canónico colectivo, ya que pretende sacar a la luz textos de diversa naturaleza que han sido relegados a un segundo plano por distintos motivos que se mencionarán más abajo. Junto a autores arrinconados por ser considerados ideológicamente ‘raros’ –muchos de los cuales tuvieron que sufrir el exilio–, aparece una larga lista de escritores, muy difundidos, de gran popularidad en la narrativa breve y en el teatro, que ingresan en el gremio de los ‘olvidados’ simplemente porque su estética había caducado en 1939. Cecilio Alonso (2008) demuestra cómo estos ‘olvidados’ popularizaron modelos naturalistas y modernistas gracias a las series semanales de novelas cortas o a otros proyectos editoriales.

Ni siquiera tras la muerte de Franco, en 1975, los autores ‘raros’ volvieron a ser publicados y la mayor parte de ellos cayó en olvido. Las editoriales apostaron por el ímpetu expresivo de los jóvenes que venían luchando con fuerza por la libertad democrática. Es complicado resolver los problemas de “discontinuidad cultural” de los que habla Vicente Llorens (2003: 94-106). El tiempo pasa para todos y van cambiando los paradigmas culturales. La era digital ha manifestado su interés por compilar los legados de la historia cultural. Después de la digitalización masiva de documentos y la creación de fondos nacionales digitalizados, llega el momento para crear las bibliotecas especializadas que ayuden a la investigación. *Mnemosine* pretende llenar este vacío bibliográfico sacando a la luz esos documentos y autores olvidados con el fin de llevar a cabo una revisión crítica de todo el período.

REPERTORIO DE AUTORES Y TEXTOS ‘RAROS Y OLVIDADOS’

Mnemosine, biblioteca digital de textos literarios raros y olvidados (1868-1939) es una biblioteca, como su nombre indica de ‘textos’ con toda la amplitud con la

⁵ Destaco sólo algunas publicaciones. Este apartado necesitaría un estado de la cuestión exhaustivo que pusiera en evidencia la gran labor realizada por investigadores nacionales e internacionales sobre figuras y géneros raros y olvidados de la Edad de Plata.

⁶ Véase en la web la colección sobre Literatura Breve publicada en la Editorial CSIC.

que ese término queda definido en el diccionario de la RAE: "Conjunto coherente de enunciados tanto orales como escritos". Se localizan dichos 'textos' con más facilidad si seleccionamos los nombres propios de sus autores o de las revistas o programas de radio en que colaboran⁷. Por eso es preciso acotar el repertorio de autores raros y olvidados. Pero seleccionar un inventario de autores es, en cierto modo, canonizarlos, pasar a formar parte de la lista de los esta vez elegidos. Mi punto de partida han sido aquellos autores citados en la bibliografía sobre raros y olvidados de la época. He ido compilándolos dentro de una base de datos que ha ido aumentando día a día y que parece no tener fin. El material utilizado para fijar el repertorio forma parte del "Material Bibliográfico de Trabajo Interno"⁸. Los repertorios que conforman el corpus inicial de autores es limitado y susceptible a ser ampliado en sucesivas fases de esta investigación. La separación entre hombres y mujeres escritores es fundamental a la hora de hacer planteamientos teóricos sobre la actividad intelectual de la época. De hecho, una de las dificultades que nos encontramos al realizar estudios sobre la actividad intelectual de las mujeres es precisamente que las bases de datos de las bibliotecas con fondos digitalizados no discriminan los géneros de los autores⁹.

POSIBLES CAUSAS DE EXCLUSIÓN

¿Cómo poner orden a ese conjunto de 'raros y olvidados'? Sainz de Robles intenta dotarles de una legitimidad histórica, pero dicha categorización es sólo un punto de partida. Los olvidos históricos son justificables pero no deseables. Tras elaborar la larga lista de autores y autoras e ir escudriñando en sus vidas y sus obras es inevitable cuestionarse sobre los motivos que pudieron influir en la exclusión de algunos autores que, sin embargo, gozaron de reconocimiento en la época. Se pueden aducir algunas causas de exclusión con el objetivo de observar tendencias que den algo de luz al oscuro fondo del conjunto homogéneo de nombres. Se observan tres grandes tendencias de exclusión dentro de cada cual se pueden observar también matices: A) Por motivos biológicos; B) Por motivos ideológicos; C) Por motivos estéticos; D) Por motivos genéricos (ser mujer y literata).

A) Motivos biológicos

Dentro de los motivos biológicos se observan dos grupos: 1) la muerte prematura del autor¹⁰; 2) su nacimiento fuera de nuestras fronteras, por lo que se le considera extranjero¹¹.

⁷ Haría falta completar esta información con un estudio sobre qué se considera 'texto raro' desde el punto de vista de la biblioteconomía y la documentación.

⁸ Se han manejado bibliografías generales de historiografía literaria, catálogos de bibliotecas, hemerotecas, editoriales, diarios, revistas y colecciones.

⁹ Debo agradecer a Kristy Hooper el listado de "Mujeres intelectuales" que ofrece en su página web, que ha sido de gran ayuda para fijar el repertorio de mujeres olvidadas.

¹⁰ Sería el caso de Josefa Vidal (Correa: 2001), que murió muy joven, pero que ha legado libros de poemas como *Vibraciones* (1902), *Cosas que pasan* (1906) y *Lira andaluza* (1906-1907), que podrían ser considerados modernistas.

¹¹ Son varios los factores por los que Alberto Galt Escobar [Alberto Insúa] pudo quedar excluido del canon. Nació en La Habana en el seno de una familia de origen español que se vio obligada a regresar a la patria tras el Desastre del 98. Como narrador y periodista participó

B) Motivos ideológicos

La lucha por conseguir la modernización del país hace que durante todo el período comprendido entre 1868 y 1936 se haga patente el enfrentamiento entre las ideas tradicionales conservadoras y las ideas más radicales e innovadoras. Dicho enfrentamiento culminará con la Guerra Civil. Después se puso en marcha el aparato ideológico de la posguerra, y el nacionalsindicalismo falangista comenzó a rehacer la historia buscando sus bases en el catolicismo, en la unidad y en la gloria histórica de España. Ese proceso de reconstrucción de un país arruinado y doliente suponía el olvido de muchos nombres propios con ideales progresista, republicanos o radicales. Se pueden percibir hasta tres grupos de escritores olvidados por el nuevo Régimen: 1.-Autores asesinados o muertos en prisión durante la contienda¹²; 2.-Entre los autores exiliados y autoexilios se encuentra una larga lista que ha recibido bastante atención crítica una vez restablecida la democracia¹³; 3.-Otro grupo de escritores que permanecieron silenciados en el país por miedo o por falta de oportunidades¹⁴.

C) Por motivos estéticos

La construcción de una literatura nacional está presidida por la evolución de las ideas y formas estéticas. La innovación puede resultar un factor tan adverso para el canon como el excesivo conservadurismo. El período histórico comprendido entre 1868 a 1936 está dominado por una pluralidad de estéticas: postromanticismo, realismo, naturalismo, modernismo, movimientos de vanguardia y compromiso social. Además existe una dinámica estética que afecta decisivamente al canon literario: las clases cultas defienden una literatura minoritaria, selecta, preocupada más por la forma y el estilo, es la alta cultura. Las clases populares serán arrastradas en masa por géneros donde la capacidad de diversión y motivación es lo fundamental. Entre los motivos estéticos podemos destacar diversas tendencias que han quedado olvidadas:

en las colecciones de novela corta, de impronta naturalista y erótica. Tras la Guerra emigró a Buenos Aires y allí murió en 1963. Álvaro Retana Ramírez de Arellano nació por casualidad en Filipinas en 1890. Obtuvo el éxito gracias a sus divertidas novelas donde refleja una sociedad matritense de buen vivir. Su fama de libertino y homosexual le llevó a conocer la cárcel y a ser represaliado con posterioridad.

¹² Algunos ejemplos son: Manuel Ciges Aparicio, mano derecha del Manuel Azaña, muere fusilado por las tropas nacionales en 1936. Fernando Mora muere víctima de un tiroteo en Zaragoza en 1936. Antonio de Hoyos y Vinent colabora en las colecciones populares de novelas cortas. Su estilo es decadente, hace explícita la homosexualidad en sus novelas. Debido a su militancia como anarquista fue apresado tras la Guerra Civil y falleció en la cárcel. Manuel Bueno Bengoechea, periodista y crítico audaz, es además autor de novelas cortas y de viajes. Apoyó la dictadura de Miguel Primo de Rivera y más tarde la sublevación nacional. Fue apresado y asesinado por milicianos en 1936.

¹³ Se encuentran algunos autores exiliados relativamente conocidos pero nunca definitivamente canonizados. Sería el caso de Ramón Gómez de la Serna. En cambio no han corrido la misma suerte otros exiliados que siguen aún en su limbo particular: José Alfonso Vidal Planas es anticlerical, bohemio y perteneciente a la CNT, se marchó a Nueva York cuando aún la Guerra no había terminado. Andrés García de Barga [Corpus Barga], escritor activo y republicano, se exilió, junto a Antonio Machado, en 1939. Posteriormente emigra a Perú donde fallecerá en 1975.

¹⁴ Diego San José de la Torre después de la Guerra no volverá a publicar nada. Tras su muerte, acaecida en Rodondela en 1962, se publica *De cárcel en cárcel* un interesante testimonio de su estancia en prisión que ha vuelto a ser reeditado en 1988.

1. Los géneros populares entre las clases medias y bajas: 1.1.–El género más popular a finales del siglo XIX es un género mixto que mezcla la dramaturgia con la música¹⁵; 1.2.–Otro de los géneros que ha quedado relegado en el canon son los cuentos, las novelas cortas y las leyendas considerados de segunda fila quizá por su extensión y porque el siglo XIX popularizó en dos sentidos: lo convirtió en narraciones para educar a los infantes, o, los difundió con literatura de prensa diaria que los grandes escritores cultivaban para ganarse un dinerillo extra¹⁶.
2. Los textos de estética naturalista que cundieron y se popularizaron después de éxito del realismo¹⁷.
3. Los textos de tema erótico¹⁸.
4. Los textos de tema ocultista¹⁹.
5. La voluntad canónica y elitista de la historiografía literaria ha desdeñado los textos que cultivan el simbolismo o el modernismo más allá de los primeros años del siglo XX²⁰.
6. La voluntad canónica y elitista de la historiografía literaria ha menospreciado a los autores pertenecientes a la bohemia finisecular²¹.

D) Por motivos genéricos (ser mujer y literata)

La crítica especializada, deseosa de trazar una tradición literaria femenina, ha ido profundizando en los textos y el valor social de escritoras que, aún no siendo canónicas, sí son conocidas. Pero los estudios sobre la emancipación literaria de la mujer escritora en España está aún por hacer y para ello necesitamos nuevos enfoques críticos y teóricos. Se han reconocido a escritoras como Cecilia Böhl

¹⁵ Aparecen las zarzuelas y las comedias musicales. El género dramático se enriquece con fórmulas alternativas como el teatro por horas y otras, más tradicionales, cobran nueva vida, como la farsa, el sainete o los entremeses.

¹⁶ Aquí habría que incluir aquellos escritores que colaboraron en el denominado "Cuento Semanal" (1907-1912): Pedro Mata Domínguez, E. Gómez Carrillo, E. Zamacois, R. Gómez de la Serna, R. Pérez de Ayala, A. de Hoyos y Vinent, L. Ruiz Contreras, José Francés, C. de Burgos, Cristóbal de Castro, José María Carretero, Álvaro Retana, Alfonso Vidal y Planas, Fernando Mora, Guillermo Díaz-Caneja, Vicente Díaz de Tejada...

¹⁷ Ha quedado olvidado el teatro naturalista que plantea conflictos morales y sociales de Manuel Linares-Rivas y Astray-Caneda, la novela costumbrista cargada de crítica social de Francisco Camba, Wenceslao Fernández Flórez y las novelas largas o cortas de Roberto Molina Espinosa.

¹⁸ El caso más notable sería el de Felipe Trigo, quien después de ejercer durante años como médico rural, abandona su profesión por el éxito de sus novelas y pasa a disfrutar de una vida placentera entre el campo y la ciudad.

¹⁹ Mario Roso de Luna es hoy en día casi un desconocido y gracias a él se difundieron las doctrinas de Madame Blavatsky en España y sin él no podríamos entender, por ejemplo, al Valle-Inclán de *Lámpara maravillosa*.

²⁰ Como por ejemplo los de Francisco Villaespesa Martín, muerto por causa natural en 1936 y que fue poeta, dramaturgo y novelista prolijo adscrito a la estética modernista durante toda su vida.

²¹ Los jóvenes con inquietudes intelectuales en la España finisecular, rebeldes y deslumbrados por todo cuanto venía de Europa, adularon la bohemia que Rubén Darío les pintó con detalle en *Los Raros*. Bohemios fueron: Armando Buscarini García Barrios, Rafael Mesa y López, Eliodoro Puche, Alejandro Sawa, Pedro Luis de Gálvez, José Nakens, Emilio Carrere Moreno, José Alfonso Vidal y Plana, Eugenio Muñoz Díaz, Antonio Rey Moliné [Dorio de Gádex], Ernesto Bark y Eduardo Zamacois.

de Faber 'Fernán Caballero', Rosalía de Castro, Emilia Pardo Bazán, Carmen de Burgos, María Martínez Sierra, Federica Montseny, Clara Campoamor, Ernestina Champourcin y Rosa Chacel... El repertorio de escritoras propuesto en la base de datos es un todo que debe ser segmentado para iniciar la investigación y elevar a categoría histórico literaria la producción intelectual de la mujer desde 1868-1936. Podemos proponer algunas causas por las que las mujeres han sido apartadas de los cánones literarios: 1.-Por motivos educativos²²; 2.-Por cuestiones sociológicas. En la época estaba mal visto que las mujeres escribieran y airearan sus ideales. Algunas de ellas utilizaron el seudónimo para dar a conocer su obra literaria (Romero: 2011); 3.-Por cuestiones estéticas: Las ideas liberales del romanticismo proporcionan una libertad y autonomía a la mujer que poco a poco irán calando en su educación. El primer paso que da la mujer es que se le permita leer por distracción y, de ahí, se les da acceso a la prensa dedicada a ellas y de manera incipiente comienza a colaborar con notas anónima, artículo, poemas y cuentos²³; 4.-Por cuestiones ideológicas²⁴; 5.-La voluntad canónica y elitista de la historiografía literaria ha arrinconado los textos escritos por mujeres porque las fuentes bibliográficas sobre autoras particulares están dispersas y son muy incompletas²⁵.

PROCEDENCIA DE LOS TEXTOS DIGITALIZADOS

La búsqueda de textos ya digitalizados con los que nutrir *Mnemosine* se llevó a cabo a través de dos instituciones que cuentan con fondos para la digitalización masiva de contenidos.

1. El primer esbozo de *Mnemosine* estuvo localizado en el servidor de la Biblioteca Digital HathiTrust que es una sociedad comercial en la que participan bibliotecas e instituciones dedicadas a la investigación y cuya misión es preservar y hacer accesible durante mucho tiempo los registros culturales. Es una sociedad en la que participan 66 instituciones de todo el mundo incluida la estadounidense Biblioteca del Congreso. El número de volúmenes digitalizados sobrepasa los 10 millones, de los que cerca de 3 millones están en dominio público. El *software* de esta biblioteca digital permite crear colecciones particulares de los usuarios. Se generó una colección propia, de dominio público, que se denomina: *Mnemosine, biblioteca digital de textos literarios raros y olvidados (1868-1939)*. En la

²² De 1866 data "Carta a Eduarda" de Rosalía de Castro, en ella se narra la situación extrema que sufren las que socialmente son conocidas como "literatas".

²³ Semanarios como *La Psiquis* (1840), *El Defensor del Bello Sexo* (1845), *La Gaceta de las Mujeres* (1845), *Álbum de señoritas* (1852) publican textos escritos por mujeres que a otras damas les era grato leer.

²⁴ Algunas mujeres como Clara Campoamor, Victoria Kent, Margarita Nelken, María Pérez Enciso, Marina Romero o María Martínez Sierra hubieran pasado a formar parte de la historia de la cultura española si los acontecimientos hubiera sido otros, si no hubieran tenido que *exiliarse o permanecer ocultas y calladas en su país*.

²⁵ Cada vez contamos con más estudios individuales sobre escritoras intelectuales de la época como los de Susan Kirkpatrick (1989, 2003) que aboga por la necesidad de abrir los estudios de fin de siglo a los estudios culturales, pero en su libro la crítica sigue tomando como ejemplo a las escritoras ya conocidas (Gómez de Avellaneda, Carolina Coronado y Cecilia Boehl). Queda mucha investigación por desarrollar. *Mnemosine* pretende que los textos de estas autoras sean visibles y que la investigación los tenga en cuenta.

búsqueda de textos se fueron incluyendo los nombres, apellidos o seudónimos de los autores seleccionados y se localizaron un total de 2.873 *ítems* que se corresponden con libros publicados por los autores considerados raros y olvidados. *HathiTrust* permite: 1) realizar búsqueda en la colección (y en los textos) de cualquier término, 2) ordenar los *ítems* por autor principal, autor secundario, lengua, año de publicación, editor, título y fecha, 3) ir al registro del *ítem* en el catálogo general, 4) acceder a él con 'vista plena', 5) visualizar el documento en distintos formatos y tamaños, 6) acceder al documento en modo imagen y también en modo texto y 7) guardar el registro o exportarlo a otra base de datos a través de *Endnote*.

Todo esto es posible si eres socio de una biblioteca que ha digitalizado²⁶ y fondos a través de Google Books, dado que el principal surtidor de textos de *HathiTrust* es precisamente esta compañía. Los usuarios de la UCM, como miembros de *HathiTrust*, tienen la posibilidad de identificarse con las claves de su correo electrónico y acceder a opciones específicas como, por ejemplo, la descarga de obras en *.pdf*, crear subconjuntos dentro de la colección general y hacerlos públicos y consultar los metadatos incluidos en las obras digitalizadas. La Universidad Complutense se convirtió en noviembre de 2010 en socio de *HathiTrust* y cuenta con más de 100.000 libros digitalizados, libres de derechos de autor fundamentalmente procedentes del proyecto BUC-Google así como de otros fondos digitalizados localmente. En el futuro próximo la Universidad Complutense de Madrid deberá negociar con dicho consorcio el acceso para estudio e investigación a documentos siguiendo las leyes española. En España quedan libre de derechos de autor los documentos cuyo autor haya fallecido hace 70 u 80 años (dependiendo de la fecha de nacimiento) y siempre que se trate de nuevas ediciones que estén protegidas por el Texto Refundido de la Ley de Propiedad Intelectual (Ley 23/2006, de 7 de julio), que está hoy vigente en España. Sería deseable que la institución universitaria española pudiera llegar a acuerdos para ofrecer contenidos de alto valor en la investigación aunque estén sujetos a derechos de autor.

2. Protocolo de entendimiento con la Biblioteca Digital Hispánica, que se llevó a cabo en la primavera de 2012 con el fin de hacer todo lo posible por digitalizar libros pertenecientes a autores 'raros y olvidados' que la Biblioteca Nacional tiene en sus fondos. En un primer acercamiento a sus fondos la Biblioteca podrá digitalizar 2.448 obras de autores 'raros y olvidados' y un total de 1.017 títulos pertenecientes a las denominadas autoras 'raras y olvidadas'. Los metadatos de dichos volúmenes cuando estén digitalizados, así como la URL de su localización pasarán a formar parte *Mnemosine* siguiendo los estándares y protocolos internacionales. Las colecciones de la BDH se han reorganizado conforme a la Clasificación Decimal Universal (CDU). Se prevé que la digitalización de los textos seleccionados pueda llevarse a cabo durante 2013 y 2014.

²⁶ En estos momentos la Biblioteca Universitaria Complutense cuenta con un nuevo proyecto en el que se digitalizarán libros publicados a partir de 1870 y hasta la fecha en la que se lo permita la legislación española sobre el copyright. Esta nueva fase de digitalización nutrirá con nuevos volúmenes nuestra biblioteca porque coincide en años con nuestro campo de estudio.

DISEÑO Y GESTACIÓN DE UNA SMARTLIBRARY

Con fecha 5/09/2012 la Subdirección General de Proyectos de Investigación del Ministerio de Economía y Competitividad concede a varios de los investigadores integrados en los grupos de investigación LEETHI (Literaturas Española y Europeas del Texto al Hipermedia) y TGEP (Temas y Géneros de la Edad de Plata) Proyecto de Investigación Nacional (Referencia FFI2012-34666) con el título de "Escritorios Electrónicos para las Literaturas-2". La concesión de dicho proyecto permitirá proseguir con la creación de la biblioteca digital Mnemosine. Lo que pretende este proyecto es no sólo la gestación de una base de datos ajustada a protocolos internacionales. Mnemosine pretende constituirse en una *Smartlibrary*²⁷, es decir, una biblioteca para la docencia y la investigación en la que se pueden implementar aplicaciones de uso personal o social y con avanzado nivel de conectividad. Se describen a continuación brevemente dos de las fases previstas de desarrollo de la biblioteca.

En la Fase II *Mnemosine* se convertirá en un repositorio institucional gracias a ODA 2.0²⁸ que es un Software científico que sirve para descubrir el patrimonio académico y que ha sido desarrollado por el grupo de investigación de la UCM dirigido por la Dra. Fernández Pampillón (2012). Uno de los objetivos de la biblioteca digital es crear sinergias de colaboración con otros grupos con el fin de crear colaborativamente museos y colecciones de objetos digitales por parte de la colectividad académica y con mínimos conocimientos informáticos y recursos tecnológicos. Aunque toda la información de la biblioteca es conveniente que se encuentre debidamente centralizada en un único servidor institucional, cada una de las fichas o entradas de las obras y autores pueden estar descentralizadas y podrán ser cumplimentadas por cualquiera de los investigadores implicados en sus propios centros de investigación. Se reconocerá científicamente la autoría individual de cada una de las fichas a través de licencias de software estandarizadas (*Creative Commons*). Así se pretende dar calidad a la biblioteca, al responsabilizarse un especialista de un determinado número de fichas, y difusión a su contenido a través de los investigadores implicados. Por ello se necesita de una amplia participación de investigadores de solvencia científica que vaya incorporando información a la red, respetando sus bases científicas. Esta forma de trabajo le dará estabilidad, carácter científico y difusión a largo plazo. En este momento la búsqueda se realiza por texto libre, pero en el futuro, se prevé que se puedan realizar búsquedas cruzadas de documentos. El acceso a la base de datos será publicado cuando el volumen de información acumulado sea científicamente relevante. Se estudiará la posibilidad de realizar un archivo con las consultas de cada uno de los usuarios. En esa Fase II del proyecto la base de datos pretende inventariar un corpus de obras autores y géneros que irá enriqueciéndose a medida que la investigación avance. Cada registro deberá tener información esencial sobre el autor y el género. Cada registro tendrá valor catalográfico y llevará la información crítica oportuna tanto de la época estudiada como de la actual. Se irán agrupando los testimonios y abriendo enlaces para recoger

²⁷ Este concepto inglés es el que mejor traduce la idea que aquí se expone de biblioteca digital. Hasta ahora, que sepamos, nadie lo ha utilizado. La creación de este tipo de biblioteca necesitará más recursos financieros.

²⁸ Se puede acceder a información general sobre este repositorio en Internet.

toda la bibliografía que ha generado un texto. Cada referencia recogerá datos de la obra en sí, datos de la crítica que ha generado o los testimonios de esa obra o autor en otras obras, bibliografía secundaria específica e información sobre el autor o autores responsables de la ficha o referencia. El modelo de ficha propuesto da cuenta de varios *corpora*: Corpus de fechas, Corpus de autores (hombre), Corpus de autoras (mujeres), Corpus de obras, Corpus de géneros, Corpus de subgéneros. Esos *corpora* permitirán aumentar el rigor documental, facilitar el acceso a información concreta y proporcionarán una visión de conjunto que permita cruzar datos para el análisis crítico y la revisión comparativa de redes de relaciones sincrónicas y diacrónicas. Con ello se pretende que el investigador pueda realizar estudios avanzados sobre autores, géneros, subgéneros, temas, etc. e investigaciones complejas que permitan la interdisciplinariedad (literatura y otras artes) y la interculturalidad (literatura escrita en español y literatura en traducción).

Aún podemos avanzar una III Fase de desarrollo de la biblioteca digital *Mnemosine* que consistirá en crear materiales didácticos y de investigación. Las anotaciones se convertirán en material de docencia o investigación que a su vez formarán parte de los recursos de la biblioteca. Otra herramienta que ya hemos desarrollado en una colaboración entre el Grupo de Investigación ILSA (UCM-Informática) y el Grupo LEETHI es *@Note*²⁹. Dicha herramienta, con la que ya hemos experimentado durante dos cursos, ha permitido a nuestros alumnos la anotación crítica y colaborativa de textos en línea. La herramienta también permite exportar las anotaciones a un editor de textos. Con posterioridad el alumno (o el crítico literario) podrá modelar las ideas y crear un ensayo que podrá publicar en línea junto a la obra estudiada.

Por lo tanto, *Mnemosine* pretende convertirse en una biblioteca de referencia para el estudio de la literatura española de la Edad de Plata. Dicha biblioteca estará inserta en un repositorio en el que el usuario encontrará herramientas para la didáctica y la investigación.

BIBLIOGRAFÍA

- Alonso, Cecilio, "Sobre la categoría canónica de 'raros y olvidados'", *Anales de Literatura Española*, 20 (2008), pp. 11-38.
- Balsach, María Josep, *La otra historia del arte: heterodoxos, raros y olvidados*, Madrid, Instituto de Cultura y Fundación Mapfre, 2006.
- Cardwell, Richard, "Bohemios, raros y liliales: la complicada historia del artista finisecular", en A. Cruz Casado (ed.), *Bohemios raros y olvidados*, Córdoba, Diputación de Córdoba y Ayuntamiento de Lucena, 2006, pp. 27-66.
- Cordero Gómez, José Ignacio, *La obra literaria de Eduardo Zamacois*, Madrid, Universidad Complutense, Servicio de Publicaciones, 2008.
- Correa Ramón, Amelina, *Alejandro Sawa: Luces de bohemia*, Sevilla, Fundación Manuel Lara, 2008.

²⁹ Este proyecto se llevó a cabo gracias a dos premios internacionales *Google Awards in Digital Humanities* (2010 y 2011), <<http://googleblog.blogspot.com/2010/07/our-commitment-to-digital-humanities.html>>, para el proyecto *Collaborative Annotation of Digitalized Literary Text*. *@Note* es una herramienta aún en experimentación y de uso restringido pero su URL es: <<http://a-note.appspot.com/>>. Se puede leer el funcionamiento de dicha herramienta digital en el libro de Luis Pablo Núñez (2013).

- Cruz Casado, Antonio, *Bohemios, raros y olvidados. Actas del Congreso Internacional celebrado en Lucena*, Córdoba, Diputación de Córdoba y Ayuntamiento de Lucena, 2006.
- Diego Fernández, Juan, *Alejandro Sawa y la Santa Bohemia*, Jerez de la Frontera, EH Editores, 2009.
- Darío, Rubén, *Los raros*, Buenos Aires, Tipografía La Vasconia, 1895.
- Ena Bordonada, Ángela (ed.), *La otra Edad de Plata. Temas, géneros y creadores (1898-1936)*, Madrid, Editorial Complutense, 2013.
- Fernández-Pampillón, Ana, *Un software científico para descubrir el patrimonio académico*, <<http://www.campusmoncloa.es/es/noticias/un-software-cientifico-para-descubrir-el-patrimonio-academico/165/>> [30/09/2012].
- Galindo Alonso, María Asunción, *La novela de una hora*, Madrid, Universidad Complutense de Madrid, Servicio de Publicaciones, 2002.
- Gallardo, Bartolomé José, *Ensayo de una biblioteca española de libros raros y curiosos*, Madrid, Imprenta y Estereotipia de M. Rivadeneyra, 1863.
- Hooper, Kristy, *The Atlantis Project: Women and Words in Spain, 1890-1936*, <<http://atlantis.kirstyhooper.net/>> [30/09/2012].
- Íñiguez Barrera, María Lourdes, *El Cuento Semanal, 1907-1912: análisis y estudio de una colección de novelas cortas*, Granada, Grupo Editorial Universitario, 2005.
- Llorens, Vicente, "La discontinuidad cultural española", en Manuel Aznar Soler (ed.), *Laberintos. Revista de Estudios sobre los exilios culturales españoles*, 2 (2003), pp. 94-106.
- Magnien, Brigitte, *Ideología y texto en El Cuento Semanal: 1907-1912*, Madrid, Ediciones de la Torre, 1986.
- Mainer, José Carlos, *La edad de plata en la literatura española. Ensayo de interpretación de un proceso cultural*, Barcelona, Asenet, 1975.
- Núñez Rey, Concepción, *Carmen de Burgos "Colombine" en la Edad de Plata de la literatura española*, Sevilla, Fundación José Manuel Lara, 2005.
- Núñez, Luis Pablo (ed.), *Escritorios electrónicos para las literaturas. Nuevas herramientas digitales para la anotación colaborativa*, Madrid, Universidad Complutense de Madrid, Servicio de Publicaciones, 2013.
- Pérez Bowie, José Antonio, *La novela teatral*, Madrid, Consejo Superior de Investigaciones Científicas, 1996.
- Pont, Jaume (ed.), *El cuento español en el siglo XIX. Autores raros y olvidados*, Lleida, Universitat de Lleida, 2000.
- Riera Guignet, Alejandro, *Emilio Carrere: el bohemio de Madrid*, Madrid, Ediciones La Librería, 2011.
- Romero López, Dolores, *Una relectura del fin de siglo en el marco de la literatura comparada. Teoría y Praxis*, Berna, Peter Lang, 1998.
- Romero López, Dolores, "La identidad velada: El uso del seudónimo en algunas literatas de la Edad de Plata", en J. Álvarez Barrientos (ed.), *Imposturas literarias españolas*, Salamanca, Ediciones Universidad Salamanca, 2011, pp. 151-170.
- Sainz de Robles, Federico Carlos, *La novela corta española: Promoción de El Cuento Semanal (1901-1920)*, Madrid, Aguilar, 1959.
- Sainz de Robles, Federico Carlos, *Raros y olvidados (la promoción de El Cuento Semanal)*, Madrid, Prensa Española, 1971.
- Sainz de Robles, Federico Carlos, *La promoción de El Cuento Semanal 1907-1925 (Un interesante e imprescindible capítulo de la historia de la novela española)*, Madrid, Espasa-Calpe, 1975.
- Serís, Homero, *Nuevo ensayo de una Biblioteca Española de libros raros y curiosos*, Valencia, Artes Gráficas Soler, 1964.



RESUMEN

El principal objetivo de la *Biblioteca Digital Mnemosine de Textos Literarios Raros y Olvidados (1868-1936)* es seleccionar, catalogar y hacer visibles en formato digital textos que pertenecen a un repertorio olvidado y que permitirá la revisión historiográfica de lo que la Dra. Ángela Ena denomina "la otra Edad de Plata" (2013). Los conceptos 'raro', 'curioso', 'olvidado' no son extraños a la historiografía literaria, pero, que sepamos, esta es la primera vez que se aplica a un repertorio de textos digitalizados y diverso en géneros literarios y formatos (libros, periódicos y revistas). Las labores que hemos realizado hasta el momento para la gestación de esta biblioteca constituyen sólo la Fase I y son las que detallamos en este capítulo: A) Localización de 247 autores hombres y 367 autoras cuyas obras caben dentro de estas categorías; B) Localización en HathiTrust de 3.876 textos ya digitalizados; C) Acuerdo de colaboración con la citada biblioteca digital estadounidense para que ponga en abierto textos que, al ser patrimonio español, deben seguir las leyes nacionales de *copyright*; D) Acuerdo de colaboración con la Biblioteca Nacional de España para la digitalización de más textos que están en sus fondos y E) Esbozo del diseño de los metadatos necesarios teniendo en cuenta los estándares internacionales. En las fases II y III se conseguirá que esta biblioteca alcance el objetivo de convertirse en una *smartlibrary*.

Palabras clave: Mnemosine, Biblioteca Digital para la Docencia y la Investigación, Raros y Olvidados, Edad de Plata, Literatura Española en Edad de Plata.

ABSTRACT

The main objective of the *Mnemosyne Digital Library of Rare and Forgotten Literary Texts (1868-1936)* is to select, catalogue and make visible in digital format texts belonging to a forgotten repertoire in order to allow the historiographical review of what Dr. Ángela Ena calls "the Other Silver Age" (2013). The concepts 'rare', 'curious' or 'forgotten' are familiar to the literary historiography, but as far as we know, this is the first time that are applied to a repertoire of texts digitized, and with diverse genres and formats (books, newspapers and magazines). The task we have already completed is the Stage I, which is described in this chapter: A) Location of 247 male writers and 367 female writers whose works fit within these categories. B) Location in HathiTrust Digital Library of 3,876 works already digitised. C) Agreement of understanding with HathiTrust to have all those books in public access because they are Spanish heritage and should follow the Spanish copyright laws. D) Agreement of understanding with the Spanish National Library for the digitization of texts of the literary period that are in their funds. E) Outline of the metadata, taking into account the international standards. In the stages II and III this digital library will reach the goal of becoming a 'smartlibrary'.

Keywords: Mnemosyne, Research Digital Library for Teaching and Researching, Rare and Forgotten Texts, Spanish Literature in the Silver Age.

Poéticas Digitales: el texto como imagen de una obra cinética

Laura Sánchez Gómez
Universidad Complutense de Madrid
laurasanchezgomezg@gmail.com

INTRODUCCIÓN

Este artículo se centrará en analizar la creación literaria en el medio digital a través del análisis y el estudio de las creaciones digitales en las cuales el medio, no es sólo usado como soporte de difusión, ordenación, distribución, almacenamiento, conversión y visualización de datos, sino como un instrumento y un lenguaje de creación. Para ello se analizará la obra de Bruno Nadeau y Jason Lewis *Still Standing* de 2005¹ en la cual el uso del medio digital implica la concepción de un lenguaje artístico y literario. Se intentará abarcar así la problemática de la llamada "textualidad" de las poéticas o literaturas digitales, procurando con esto llegar a comprender cómo esta "textualidad" no radica en la analogía con el texto sino en su propia acepción como lenguaje. Esto significaría que la presencia o en ocasiones la no-presencia del texto no son los vehículos únicos de significación en la obra. La literariedad de las llamadas literaturas digitales no recae en este caso exclusivamente en el texto sino en las relaciones que se establecen entre sus características formales y estéticas tales como, la materialidad de la obra digital (texto/imagen), su intrínseco carácter presencial o de presencia, su esencia performativa y su problemática cinética o inmóvil. Para abordar estas cuatro cuestiones se recurrirá a la continua comparación de la obra *Still Standing* con otras obras que proponemos como sus antecedentes analógicos. Estas obras se caracterizan por tratar temas fundamentales como la interacción, la multimedialidad o el cinetismo.

¿QUÉ QUEREMOS DECIR CUANDO NOS REFERIMOS A POÉTICAS DIGITALES?

¿Hay significativa experimentación –generadora de algún conocimiento– ahí donde se recurre a lenguajes nuevos, o allí donde esa recurrencia se encuentra sustentada por una reflexión sobre la práctica artística y sus propios materiales, dando lugar, entonces, a la configuración de *metapoéticas tecnológicas*? (Ré, 2011: 22).

¹ *Still Standing* es una instalación interactiva de Bruno Nadeau y Jason Lewis de 2005 que se puede encontrar en el segundo volumen de la colección de *Electronic Literature Collection*, <http://collection.eliterature.org/2/works/nadeau_stillstanding.html>.

Cuando hablamos de literatura digital casi siempre los límites son difusos y no es tanto un problema de invisibilidad de límites como de carencia de estos. La denominación de "literaturas digitales" queda demasiado amplia quizás por pretender acaparar y comprender muchas prácticas de escritura digital y literatura en pantalla que no por su forma, o su interfaz, ni siquiera por su contenido, pero sí por la relación entre ambos, no inciden en una retórica de autodefinición y auto consideración que nos parece muy necesaria. Para explicar esto propongo utilizar un término de Rita Raley que es el concepto de "objeto crítico"². Esto significaría que el "objeto" (la literatura en este caso) debe y tiene que problematizar su relación entre la forma y el fondo. Es decir, el objeto ha de ser auto-reflexivo. Este concepto, punto clave de partida, es desarrollado por Rita Raley alegando que ya no es necesaria esa defensa encarecida de lo digital porque lo digital ya está consolidado, de manera que las obras también han de tener un valor literario y un valor poético que justifique su forma no analógica. Cuando hablamos de poéticas digitales, entendemos por tanto que éstas asuman un entorno digital, pero siempre que lo utilicen como medio, canal y soporte (artísticamente entendido como lienzo), a la vez que como objeto de reflexión y problematización. Es en definitiva una retórica que juega con la autodefinición y la reconsideración de las prácticas creativas, sus modos, sus formas y sus medios. Precisamente aquí es donde se encontraría esa auto-reflexividad³ de la que hablábamos anteriormente.

La obra de Bruno Nadeau y Jason Lewis *Still Standing* de 2005 (Fig. 1) es una instalación textual digital, interactiva, multimedial y cinética. En esta obra mientras el espectador o lector se mantiene en movimiento frente a la pantalla, el texto permanece ilegible, fluido y fragmentado. Es un texto desfallecido, caído al suelo, las letras parecen hojas secas movidas por el viento o por el paso del espectador. Cuando el lector o espectador se detiene frente a la pantalla en una postura de lectura, el texto deviene comprensible: se levanta y va componiendo la sombra del espectador a la vez que sus letras, ahora sí legibles, forman palabras y las palabras forman frases y significados. *Still Standing* es una obra que "trabaja asumiendo en forma más o menos explícita...el espacio tecnológico de época: hacia él, contra él, con él" (Kozak, 2011: 54). De esta manera y a través de su carácter digital, se replantean y se cuestionan los procesos de lectura asociados al medio digital. Es decir: se cuestiona el medio a través de sus propias posibilidades técnicas y estéticas. En esta obra se pone de relevancia lo que explica Roberto Simanowski: "By definition, digital literature must go beyond what could be done without digital media. By definition, digital literature must be more than just literature otherwise it is only literature in digital media" (2009: 12). De esta manera *Still Standing* no podría ser impresa, es decir sólo puede "ser" a través del medio digital y tampoco podría "ser" sin la participación del lector/espectador que le da forma,

² Se hace referencia aquí al concepto de "critical object" mencionado por Raley en su artículo *Living Letterforms: The Ecological Turn in Contemporary Digital Poetics*, 2001, p. 888.

³ Esta idea de auto-reflexividad de la obra digital ha sido matizada también por otros teóricos, considerándola asimismo un concepto clave para la definición de una obra de arte o de una poética digital. Rescato aquí dos aproximaciones que hacen Claudia Kozak y su grupo Ludión: "al interior de las poéticas tecnológicas existen posicionamientos respecto de la relación entre arte y tecnología" (2012: 182), y Barret Watten: "What is poetics? To begin with, it is the self-reflexive mode of the 'making' of the work of art or cultural product (...) poetics questions the nature and value of the work of art (...)" (2006: 335).

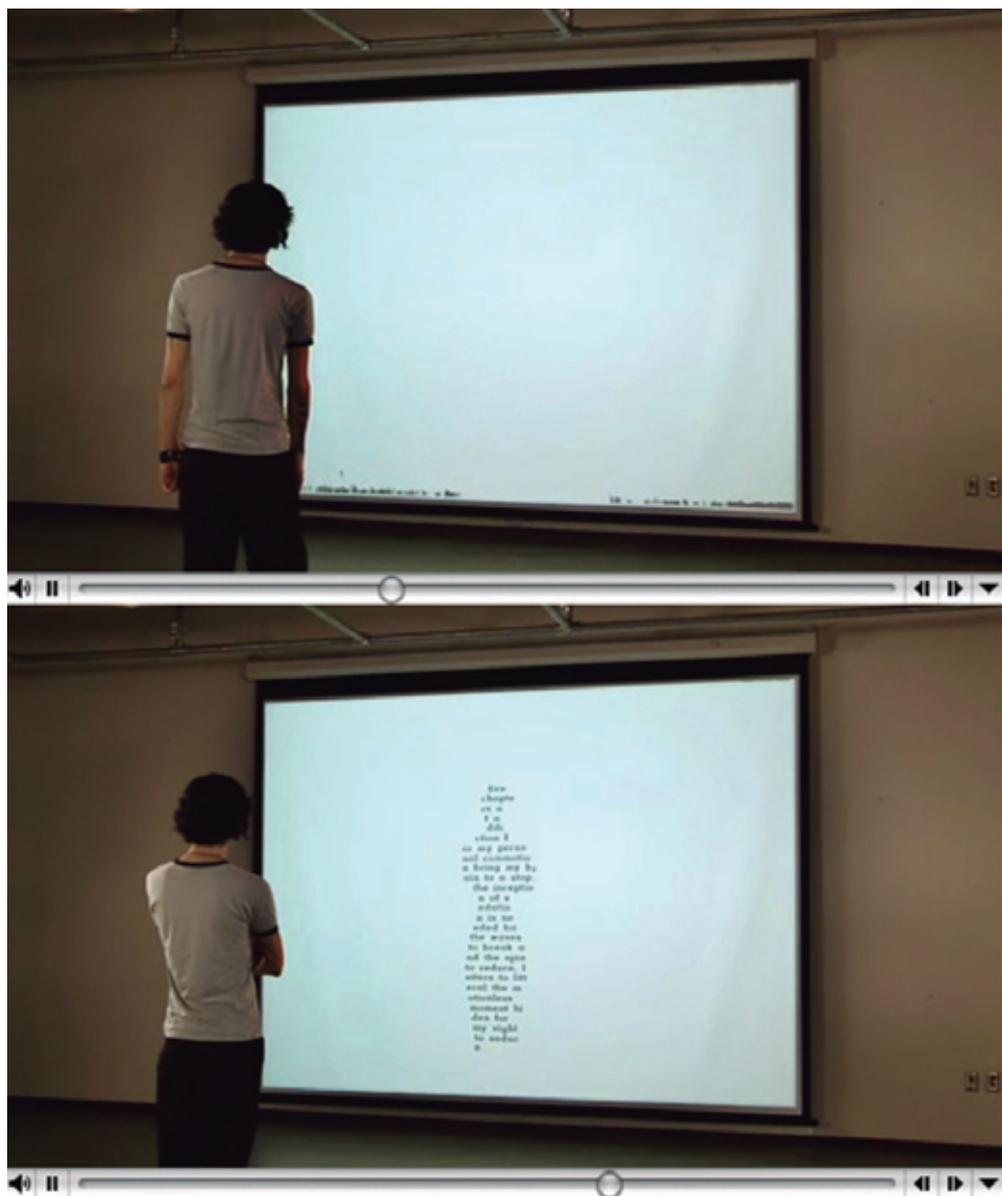


Fig. 1. Bruno Nadeau y Jason Lewis, *Still Standing*, 2005.

en este caso además, su propia forma. La obra *Still Standing* de Bruno Nadeau y Jason Lewis es por tanto, bajo nuestra consideración, un claro ejemplo de lo que entendemos como poética digital. Vemos también cómo la palabra parece haberse convertido en un componente más de la imagen que proyecta la pantalla. La palabra en esta obra es una parte compositiva más dentro del perfil de la sombra del espectador perdiendo parte de su significado textual. De esta

manera, la propia palabra ha dejado de ser "palabra" como tal, convirtiéndose en otra cosa, ¿tal vez imagen?

EL TEXTO COMO IMAGEN Y EL TEXTO PRESENCIA.

In a work that is as visual as it is literary, does it mean to speak of art, or of literature? (Ricardo, 2009: 2).

Sobre la conversión del texto en imagen a través de la pantalla se ha hablado mucho y son numerosísimas las relaciones entre ambas anteriores a las posibilidades digitales. Partiendo de la *Ekfrasis* (la descripción de una imagen en palabras), casi todos los antecedentes estéticos que se proponen en esta comunicación parten de la problematización del límite entre texto/imagen: poesía concreta, instalación, créditos cinematográficos y arte cinético.

Surge en este punto una inevitable pregunta: ¿en qué medida podemos considerar al nuevo texto digital un texto visual? ¿Es un texto contenido en una pantalla o ha pasado a ser una imagen? Para acercarnos teóricamente a este punto propongo dos investigaciones recientes y dos sistemas analíticos y categóricos del llamado texto/imagen. El primero sería el libro de Liliane Louvel *Poetics of Iconotext* de 2011, en el cual Louvel introduce el concepto de "iconotexto" con el que intenta expresar la fusión entre texto e imagen en un deseo de fundir dos objetos distintos en uno nuevo e inseparable. El segundo, *El Lectoespectador*, es un libro de Vicente Luis Mora (2012) en el cual, aunque muy influenciado por Louvel, no se hace referencia al objeto en sí, sino al nuevo proceso de recepción de información texto/visual por el cual el lector se convierte en una mezcla de espectador y lector al mismo tiempo.

Es obvio que la lectura hoy no se limita a la palabra impresa, el texto/imagen abunda en nuestros días inundando nuestra rutina y nuestras retinas, metiéndose en nuestros hogares, asediándonos de mano de las pantallas. Habría que plantearse entonces de qué manera el texto puede ser vaciado de connotación literaria, de qué manera se puede acercar al icono, y en qué punto exacto dejaría de ser texto para pasar a ser otra cosa. Reflexionando sobre esta cuestión nos daremos pronto cuenta de que lo digital y la pantalla no implican aquí ninguna novedad sustancial.

Vanguardias como el letrismo ya jugaban con la sustracción del contenido literario del texto. Con los ejemplos de la poesía concreta vemos cómo el texto/imagen pierde referencialidad y se objetualiza, produciéndose una reducción del lenguaje paradójicamente para aumentar su capacidad comunicativa. El poema de Alberto Hidalgo "Jaquica" (Fig. 2) publicado en *Química del espíritu*⁴, es sólo un ejemplo de la tradición visual y concreta de países como Francia, Portugal, Brasil o Japón. Por otro lado las instalaciones textuales en el arte contemporáneo convierten el texto en un objeto real, ocupando un espacio tridimensional como puede ser una galería de arte. Vemos a artistas como Jenny Holzer y Roni Horn⁵

⁴ Nos referimos al poema de Alberto Hidalgo publicado en *Varios. 9 libros vanguardistas*, p. 39, que originalmente fue publicado en 1923 en *Química del espíritu*. Vemos cómo la composición flotante de las letras hace referencia al título "Jaquica", jugando con la ilegibilidad del texto que pasa a ser prácticamente un icono.

⁵ Jenny Holzer y Roni Horn son artistas estadounidenses que trabajan ampliamente la instalación textual en sus obras.



Fig. 3. Bruce Nauman, None Sing, Neon Sign, 1970.

espacio de la galería, *Wall Drawings* 1968-2007, como en sus instalaciones en tres dimensiones, *Maquette for One, Two, Three* de 1979, ya introduce primitivamente al espectador en un juego de participación que incorpora una dimensión temporal-experiencial en la obra de arte. Carmen Crouzeilles, explica también el arte óptico como un claro antecedente estético de las poéticas digitales y cómo su natural evolución incluye obligatoriamente el desarrollo de sus posibilidades mediante el medio digital:

El arte óptico ha encontrado grandes posibilidades de desarrollo en el cinetismo virtual generado por computadoras con programas de diseño. Por último, hay obras que solicitan el desplazamiento del espectador con el propósito de observar efectos cambiantes que dependen del punto de vista (Crouzeilles, 2012: 49).

La objetualización de la palabra de mano de la poesía concreta, la colocación y ubicación en un contexto espacial fuera de los marcos tradicionales como la página o la pantalla a través de la instalación textual y la aparición de la experiencia a través de la incorporación de una dimensión temporal en el arte óptico y el arte cinético, implican un uso preciso del tiempo y el espacio. La palabra ya es materializada por el arte contemporáneo y las vanguardias literarias, es vaciada de referencialidad y dotada de una durabilidad y una ubicuidad. Por tanto el tratamiento de la temporalidad y la espacialidad es precisamente el punto en común entre las poéticas digitales y sus referentes analógicos. El texto en las

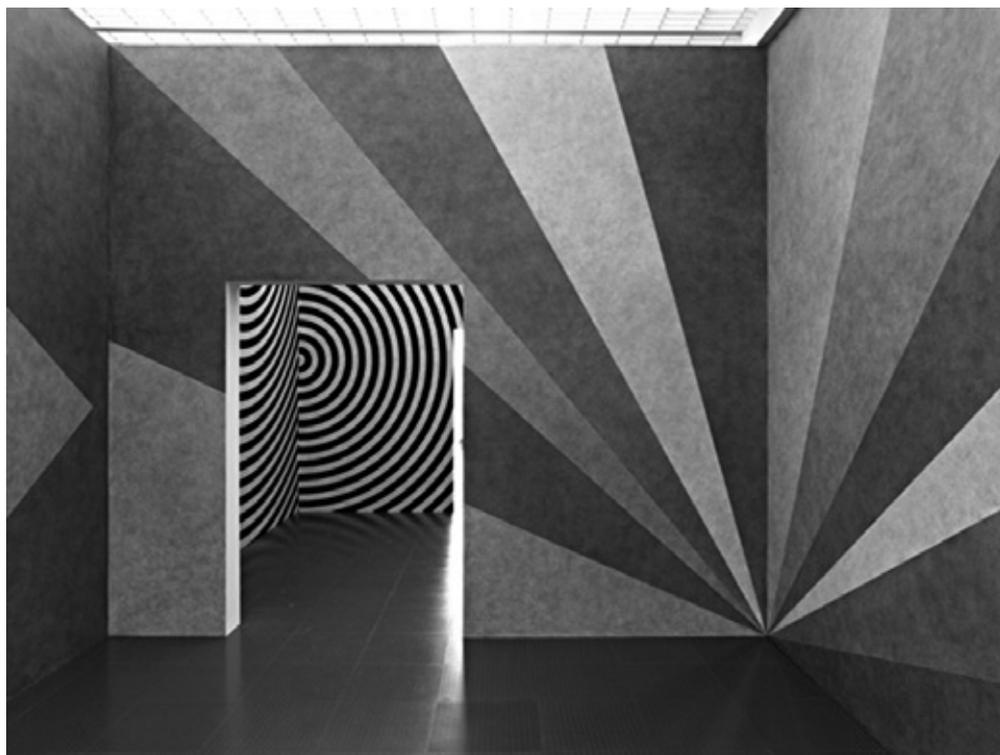


Fig. 4. *Sol* Lewitt, MASS MOCA Installation, 1968-2007.

poéticas digitales al contrario que en las páginas de un libro (en las que dura el tiempo que es leído, y la durabilidad por tanto está en el proceso de recepción), ocupa la pantalla durante un tiempo determinado que podríamos llamar su durabilidad. Así Katherine Hayles propone pensar en el poema digital “as a machine to organize time” (Hayles, 2006: 181). Esta dependencia sensitiva hacia los contextos espacial y temporal es lo que Katherine Hayles comprende que establecería las poéticas digitales como procesos, “call a digital text a process rather than an object I highlight by referring to the time of performance for an electronic text *versus* the time of production for print” (Hayles, 2006: 185). De modo que el texto ya no “está” sino que ahora “es”, y así la experimentación literaria se acerca donde ya se acercaran otras prácticas artísticas: a la experiencia. La experiencia es, como ya hicieron en sus obras Richard Long o Dennis Oppenheim⁶, la nueva identidad de la obra, la obra “es” mientras está sucediendo y lo que queda después sólo es documentación del proceso. *Still Standing* (2005) sucede y tiene lugar mientras el espectador está frente a ella, mientras la construye, así cada lectura es diferente porque no hay dos acciones

⁶ Rescatamos aquí el marcado carácter experiencial de las obras de Richard Long y Dennis Oppenheim, que se caracterizan por la intervención en el paisaje. La ejecución da como resultado muchas veces obras efímeras de las que sólo nos queda la documentación que aporta el artista del proceso.

iguales. Efectivamente hablar de producción aquí no tiene sentido y los términos "presencia" y "proceso" van adquiriendo cada vez más protagonismo.

Por tanto, que las poéticas digitales incorporen las dimensiones espacial y temporal, que el texto en ellas dure y ocupe, que sean un proceso como explica Katherine Hayles, implica que tengan una presencia. El texto en la obra *Still Standing* es presencial y ésta es una de las razones, como veremos más adelante, de que la obra sea entendida también como un proceso y no como un objeto. De esta manera podemos observar cómo las teorías de la desmaterialización del objeto de arte promulgadas por Lippard⁷ van evolucionando en lo que a creación digital se refiere. Aunque la desmaterialización aquí no hace referencia al carácter "virtual" de la obra (por estar el texto contenido en una pantalla podríamos pensar que no es material sino lumínico) sino precisamente de este paso del objeto al proceso.

EL TEXTO CINÉTICO Y EL TEXTO PERFORMATIVO

Considerando el significado de "proceso" podemos ver cómo en la segunda acepción de la Real Academia de la Lengua Española encontramos "Transcurso del tiempo". La presencia de la palabra, como ya hemos visto, ha de durar un tiempo, así la aparición de la experiencia va de la mano de la incorporación de la dimensión temporal. El texto tradicional de las páginas de un libro, que antes requería un tiempo de lectura, ahora en *Still Standing* ocupa la pantalla y se mueve en ella, posibilitando o no la lectura por parte del espectador. El texto incorpora el movimiento o el cinetismo, como ya lo hacían las obras de arte cinético.

Barbara Brownie⁸ elabora un esquema bastante interesante sobre los grados de cinetismo de la palabra o la tipografía. Es su estudio define la tipografía "serial", aunque no tenga movimiento en sí misma, como una tipografía temporal. Brownie llega en su tesis a la Tipografía fluida, "fluid typography" (2007: 8), que sería aquella que se transforma y transfigura, que cambia en sí misma. De esta manera, tendríamos aquí diferentes estados semióticos dentro de la letra que se trasforma: estados legibles y estados presemióticos (que son aquellos en los que no hay una referencialidad textual y la "letra", hasta que vuelve a ser "legible", perdería su significación textual). En *Still Standing* (2005) mientras las letras se amontonan y se superponen en "el suelo" de la pantalla, serían ilegibles; en este caso son "materia" visual pero no "textual". Podríamos entender aquí la palabra, el texto, como una simple imagen en movimiento. Pero, ¿qué ocurre cuando en una obra los estados legible y pre-semióticos se van sucediendo? En este caso

⁷ En 1937 Lucy R. Lippard escribe *Six Years: The Dematerialization of the Art Object*, donde revisa diferentes obras de arte conceptual que trascienden a la materialidad de sus soportes. Estos conceptos son seguidos y revisados también a través de la literatura digital por autores como Francisco J. Ricardo en *Literary Art in Digital Performance. Case Studies in New Media Art and Criticism* (2009).

⁸ Barbara Brownie desarrolla en varias publicaciones de las cuales aquí sólo se mencionarán dos (*One Form, Many Letters: Fluid and transient letterforms in screen-based typographic artefacts* de 2007, y *Fluid Characters in Temporal Typography* de 2012), una categorización de los diferentes tipos de tipografías según su movimiento, estaticidad y su legibilidad, llegando así a revisar el término de Eduardo Kac, "fluid" (2007: 7), para adaptarlo a la tipografía: "the identity of forms appears to change over time. At one moment they may appear typographic, but at another they may appear pictorial or abstract" (2012).

¿no sería esa concepción dialéctica del montaje surgida por confrontación (o choque para Walter Benjamin⁹) donde radicaría la verdadera explosión creativa y el desarrollo de la facultad argumentativa y discursiva de las poéticas digitales? Entran aquí en juego las relaciones entre planos, instantes y cuadros, creando diálogo, discurso y significado, enriqueciendo o dificultando quizás la visibilidad (cognitiva) y la legibilidad de las imágenes fijas o materiales que lo componen. En ocasiones también dificultando la lectura del texto cuando queremos leerlo como si siguiera anclado en la página de un libro.

En este sentido la articulación dialéctica de los títulos de crédito en cuanto a su montaje nos da muchas pistas del uso de la imagen textual cinética en la pantalla. El texto en *Still Standing* (2005), al igual que en el texto usado en el lenguaje cinematográfico, alterna estados legibles con estados pre-semióticos e introduce el movimiento en la pantalla produciendo narratividad. El lenguaje cinematográfico y las poéticas digitales no coinciden sólo en la disposición del texto en movimiento por la pantalla y en la utilización del lenguaje surgido de la confrontación de planos, sino, y lo más importante, en cómo el texto pierde parte de su función literaria para pasar a una significación más icónica. Este concepto de *shock*, de confrontación de planos o montaje dialéctico, es lo que caracteriza la creación cinematográfica soviética, donde, según explica Deleuze, "el conjunto se refleja en cada parte y cada (...) parte reproduce el conjunto (...) la imagen-movimiento es célula de montaje y no simple elemento" (1984: 57). Esto lo podemos observar en las obras de Saul Bass para Otto Preminger (Fig. 5) o Hightchcock, en las que la concepción total de la obra es representada por una de sus partes, en este caso lo títulos de crédito. Es decir, se aprende a representar el todo por las partes, de modo que una imagen texto visual, icónica y "literaria", resume la narración de una manera concisa y metafórica. De nuevo podemos rescatar aquí ideas como la concepción fractal de la poesía digital, es decir la creación directa de fragmentos o de partes sin un todo previo.

El movimiento y la utilización de diversos medios (texto e imagen que se suceden o se solapan) como canales de información, problematizan la recepción de las obras o poéticas digitales. Así *Still Standing* podría considerarse una obra multimedial. No estamos hablando de una experiencia poliestética, como definiría Molinuevo¹⁰, sino de un evento multimedial en el que no se trata del número de sentidos por los que podamos percibir y disfrutar de la obra sino del número de canales que ésta misma utiliza para llegar a su receptor. Pero ¿tiene que ver la performatividad con la cantidad de movimiento que haya al otro lado? ¿No implicaría esto también una contemplación, sólo que cambiando el objeto estático por uno dinámico? La sucesión entre imagen y texto y el movimiento de ambos dentro de la obra nos hacen pensar en una mayor "vida" dentro de ella. El cambio que supone pasar del objeto al proceso es un claro intento de acercamiento hacia el espectador, ahora más conmovido y temiendo ser alcanzado.

⁹ Aunque Walter Benjamin no deposita en el cine su plena confianza, lo analiza como imagen-movimiento relacionándolo a su vez de alguna manera con el concepto de montaje-dialéctico soviético, que más tarde desarrollaría en sus teorías sobre cine el filósofo Gilles Deleuze.

¹⁰ José Luis Molinuevo, en su conferencia "Estética política de las imágenes intencionales" en el Media-Lab del Prado en Madrid el 30 de mayo de 2013, aclara que las imágenes producen experiencias poliestéticas que van más allá de la vista [<http://medialab-prado.es/article/estetica-politica-de-las-imagenes-inintencionales>].



Fig. 5. Saul Bass, El hombre del brazo de oro, dir. Otto Preminger, 1955.

Sin embargo, ¿no debería la performatividad implicar una relación todavía mayor entre espectador y obra?

Nos preguntamos entonces si el movimiento implica performatividad ¿Tiene el texto que ocupa un espacio y un tiempo determinados que moverse para performar? ¿Es el texto cinético performativo de por sí? En este caso cabe destacar que el texto cinético de por sí (el utilizado en los títulos de crédito por ejemplo) no es performativo; a pesar de ser un texto en movimiento, el espectador no interviene más que como “espectador pasivo” recibiendo la información. En el concepto de performatividad en la poéticas digitales entran en juego las relaciones del espectador/lector con el dispositivo y las relaciones funcionales y lógicas del dispositivo que caracterizan el medio técnico. De manera que se podría hablar de nivel performativo cuando ocurre algo durante el proceso interactivo entre la obra y el espectador, cuando se desencadenan relaciones entre ambos que modifican y construyen el significado de la obra. Volviendo al ejemplo de *Still Standing* observamos que es el espectador/lector el que construye literalmente la obra, dejando atrás esas ilusiones “interactivas” en las que el lector sólo puede observar cómo la obra “sucede”. Esa actitud kantiana de contemplación de la belleza desde la ventana ha desaparecido, aquí el significado es completado por el lector a través de su interacción y la comprensión de la obra requiere un esfuerzo extra.

Como vemos, las posibilidades lingüísticas (de lenguaje, no vinculadas al texto) de las poéticas digitales son numerosísimas. Su textualidad no radica en su analogía con el texto, sino en su propia acepción como lenguaje. Por lo tanto, es como lenguaje, cuando el texto digital, presencial, performativo y móvil, pero sobre todo temporal, adquiere toda su significación y cuando sus interrelaciones (también a nivel semiótico) producen significado e idea. Las poéticas digitales han de encontrar su propio modelo de lenguaje basado en su funcionamiento

como sistema, en el que intervienen las relaciones entre el lector, la máquina, la imagen, el texto y el lenguaje.

BIBLIOGRAFÍA

- Brea, José Luis, "Algunos pensamientos sueltos acerca de arte y técnica", *Artefacto*, 6 (2007).
- Brownie, Barbara, "One Form, Many Letters: Fluid and transient letterforms in screen-based typographic artefacts", *Networking Knowledge: Journal of the MeCCSA Post-graduate Network*, 2 (2007), pp. 1-15.
- Brownie, Barbara, "Fluid Characters in Temporal Typography", *Fusion*, 1 (2012), <<http://www.fusion-journal.com/issue/001-fusion/fluid-characters-in-temporal-typography/>>.
- Crouzeilles, Carmen, "VJing: producción de experiencia, espectáculo, ilusión y montaje. Una articulación entre las tecnologías del siglo XXI y algunos rastros del arte del siglo XX", en Claudia Kozak (comp.), *Poéticas tecnológicas, transdisciplina y sociedad. Actas del Seminario Internacional Ludión/Paragraphe*, Buenos Aires, Exploratorio Ludión, 2011, pp. 89-100.
- Deleuze, Gilles, *La imagen-movimiento. Estudios sobre cine 1*, traducción de Irene Agoff, Barcelona, Paidós, 1984.
- Golding, Alan, "Language Writing, Digital Poetics, and Transitional Materialities", en Adelaide Morris y Thomas Swiss (eds.), *New Media Poetics. Contexts, Technotexts, and Theories*, London, MIT Press, 2006, pp. 249-284.
- Hayles, N Katherine, "The Time of Digital Poetry: From Object to Event", en Adelaide Morris y Thomas Swiss (eds.), *New Media Poetics. Contexts, Technotexts, and Theories*, London, MIT Press, 2006, pp. 181-210.
- Hidalgo, Alberto, "Química del espíritu", en *Varios. 9 libros vanguardistas*, Lima, El Virrey, 2001, p. 39.
- Kozak, Claudia, "Teco-poesía experimental y políticas del acontecimiento", en Claudia Kozak (comp.), *Poéticas tecnológicas, transdisciplina y sociedad. Actas del Seminario Internacional Ludión/Paragraphe*, Buenos Aires, Exploratorio Ludión, 2011, pp. 53-64.
- Lippard, Lucy R., *Seis años: la desmaterialización del objeto artístico de 1966 a 1972*, traducción de M^a Luz Rodríguez Olivares, Tres Cantos (Madrid), Akal, 2004.
- Louvel, Lilliane, *Poetics of the Iconotext*, edición de Karen Jacobs, traducción de Laurence Petit, Farnham, Ashgate, 2011.
- Mora, Vicente Luis, *El Lectoespectador*, Barcelona, Seix Barral, 2012.
- Raley, Rita, "'Living Letterforms': The Ecological Turn in Contemporary Digital Poetics", *Contemporary Literature*, 4 (2001), pp. 883-913.
- Ré, Anahí Alejandra, "Arte, ciencia, experimentación: expoésia y metapoéticas tecnológicas", en Claudia Kozak (comp.), *Poéticas tecnológicas, transdisciplina y sociedad. Actas del Seminario Internacional Ludión/Paragraphe*, Buenos Aires, Exploratorio Ludión, 2011, pp. 21-28.
- Ricardo, Francisco J., "Juncture and Form in New Media Criticism", en Francisco J. Ricardo (ed.), *Literary Art in Digital Performance. Case Studies in New Media Art and Criticism*, New York, Continuum, 2009, pp. 1-9.
- Saemmer, Alexandra, "Aesthetics of surface, ephemeral, re-enchantment and mimetic approaches in digital literature", *Neohelicon*, 36 (2009), pp. 477-488.



RESUMEN

La literatura digital se caracteriza por usar el medio tecnológico como un instrumento y un lenguaje de creación. Se pretende analizar la obra de literatura digital de Bruno

Nadeau y Jason Lewis *Still Standing* de 2005, en la cual el uso del medio digital implica la concepción de un lenguaje artístico y literario. Se intentará abarcar así la problemática de la llamada "textualidad" de las poéticas o literaturas digitales, procurando con esto llegar a comprender cómo esta "textualidad" no radica en la analogía con el texto sino en su propia acepción como lenguaje. Esto significaría que la presencia o en ocasiones la no-presencia del texto no son los vehículos únicos de significación en la obra. La literariedad de las llamadas literaturas digitales no recae en este caso exclusivamente en el texto sino en las relaciones que se establecen entre sus características formales y estéticas tales como, la materialidad de la obra digital (texto/imagen), su intrínseco carácter presencial o de presencia, su esencia performativa y su problemática cinética o inmóvil. Para abordar estas cuatro cuestiones se recurrirá a la continua comparación de la obra *Still Standing* con otras obras que proponemos como sus antecedentes analógicos. Estas obras se caracterizan por tratar temas fundamentales como la interacción, la multimedialidad o el cinetismo.

Palabras clave: Literatura digital, poéticas digitales, textualidad, performatividad, Cinetismo.

ABSTRACT

Digital literature is characterized by using the technological environment as a tool and a language of creation. It aims to analyze Bruno Nadeau and Jason Lewis's work of digital literature, *Still Standing* from 2005, in which the use of digital media involves the conception of artistic and literary language. We will try to include the problematic of "textuality" in literatures and digital poetics, trying with this to understand how this "textuality" does not lay in its analogy with text but in its own meaning as language. This would mean that the presence or sometimes the non-presence of the text are not the only vehicles of meaning in the literary work. The literariness of digital literatures do not lie only in the text but in the relationships that exist between formal and aesthetic characteristics such as the digital work materiality (text/image), its inherent presence's nature, its performative essence and its kinetic or immobile problematic. To deal with these four subjects will be used the continually comparison of *Still Standing* work with other works that are proposed as their analog antecedents. These works are characterized by treating key questions such as the interaction, multimedialism or kinetism.

Keywords: Digital Literature, digital poetics, textuality, performativity, kinetism.

La creación y gestión del archivo digital valleincliniano: *corpus manuscrito e impreso*

MARGARITA SANTOS ZAS (COORD.)

Grupo de Investigación Valle-Inclán, Universidade de Santiago de Compostela
margarita.santos@usc.es

FRANCISCA MARTÍNEZ RODRÍGUEZ

Grupo de Investigación Valle-Inclán, Universidade de Santiago de Compostela
francisca.martinez@usc.es

ROSARIO MASCATO REY

Grupo de Investigación Valle-Inclán, Universidade de Santiago de Compostela
rosario.mascato@usc.es

1. EL FONDO DOCUMENTAL VALLEINCLANIANO DEL GIVIUS: LA PUESTA EN MARCHA DE UN ARCHIVO DIGITAL

Presentar al Grupo de Investigación Valle-Inclán de la USC (GIVIUS) es paso previo para explicar su trayectoria y objetivos, con los que se relaciona la construcción y gestión de su fondo documental y la puesta en marcha de su archivo digital.

Integramos el GIVIUS (<http://www.usc.es/es/catedras/valle-inclan/givius>) tres profesores de la USC, su núcleo estable desde hace casi dos décadas, cuatro doctores –tres de ellos colaboradores externos y el cuarto con un contrato posdoctoral– y cuatro doctorandas –dos de las cuales son contratadas con cargo a proyectos–. Es decir, somos once miembros, con formación teórica y experiencia compartida a lo largo de años, que hemos tratado de impulsar, desarrollar o culminar tareas difícilmente abarcables de modo individual.

Por otra parte, a partir del año 2002 me hice cargo de la dirección de la *Cátedra Valle-Inclán* (financiada en el marco del **"Programa Universidades"** del Grupo Santander), creada con el fin de apoyar y ampliar las actividades del GIVIUS. Así, las iniciativas de la *Cátedra*, que dispone desde 2009 de un portal en la Biblioteca Virtual Miguel de Cervantes (<<http://www.cervantesvirtual.com/portal/catedravalleinclan>>), se han orientado, por una parte, a la formación de los integrantes del grupo de investigación; por otra, a la difusión de la investigación valleincliniana. En respuesta al primer objetivo nació el *Seminario Permanente*, impartido por especialistas en la obra del escritor gallego, proporciona a sus participantes la oportunidad de debatir temas y problemas relacionados con

los intereses investigadores del Grupo¹. Para alcanzar el segundo, se puso en marcha uno de nuestros proyectos más acariciados, la *Biblioteca de la Cátedra Valle-Inclán*, Servicio de Publicacións e Intercambio Científico de la USC; una colección con sello propio destinada a la difusión de los estudios y ediciones valleinclinianos, que da cabida a obras propias y ajenas². Por último, más de una década (2001-2013) llevamos impulsando el *Anuario Valle-Inclán/Anales de la Literatura Española Contemporánea* (Society of Spanish and Spanish-American Studies), que se ha convertido en un punto de referencia para los estudios valleinclinianos. Sumemos a todo ello, también como mecanismo de difusión, la organización de exposiciones (2008 y 2009) y congresos (6 hasta la fecha, el último en octubre de 2011), dedicados al escritor gallego.

En estos años hemos afrontado no pocos desafíos, obtenido algunos logros particularmente destacables y desde hace algún tiempo hemos mirado hacia el horizonte que nos ofrecen las humanidades digitales.

Nada de la tarea llevada a cabo hubiese sido posible de no haber contado con financiación: cinco proyectos consecutivos del MEC, desde 1994, en que asumí la dirección del Grupo; y dos de la Xunta de Galicia, uno de ellos dirigido por Javier Serrano, han permitido consolidar nuestras líneas de trabajo. En la actualidad disponemos de un Proyecto subvencionado por el MEC ("La obra y el legado manuscrito de Valle-Inclán: ediciones y estudios críticos", ref. FFI2011-24130), incorporado a la Red-Investiga, y de una subvención de la Xunta de Galicia, concedida a GPC en el marco de las "Axudas do Programa de Consolidación e Estruturación de Unidades de investigación Competitivas" (ref. CN 2012/097).

Nuestro trabajo, cuyos resultados se han ido publicando en diversos medios especializados (ver la web del GIVIUS arriba citada), se asienta sobre un fondo documental y bibliográfico –prensa y obra editada–, en buena parte resultante de una indagación continuada en bibliotecas y hemerotecas tanto locales y nacionales como europeas y americanas; al igual que disponemos de la bibliografía primaria del escritor anterior a 1936; y contamos con una bibliografía secundaria actualizada, que permite abordar casi cualquier tema valleincliniano desde perspectivas diversas.

A este *corpus* impreso se han sumado en noviembre de 2009 los originales del desconocido Legado Manuscrito de Valle-Inclán, depositados en la USC por voluntad de sus propietarios³. Hasta esa fecha los autógrafos valleinclinianos se contaban con los dedos de una mano, aunque se han incrementado estos últimos años con algunas donaciones (diez cartas) y adquisiciones propias (poemas, cartas y tarjetas postales). A ese reducido patrimonio original se ha agregado

¹ Con un máximo de cuatro seminarios por curso académico, cuya duración no excede las ocho horas, en este foro de debate, que inauguró sus actividades en 2003, han intervenido dieciocho profesores de universidades o centros de investigación españoles, europeos y americanos, todos ellos reconocidos valleinclinianos.

² La colección tiene dos líneas editoriales: ediciones (facsimilares, anotadas, críticas...) y monografías y desde que inició su andadura, ha publicado ocho volúmenes, cuyos autores y títulos pueden verse en la web del GIVIUS, arriba citada. Baste decir que la colección se inauguró –y en esto ha sido pionera– con la edición facsímil del manuscrito de Valle-Inclán, *Mi Bisabuelo*, ed. de Jorge Devoto (2007); y el último de los títulos editados es un volumen colectivo: *Valle-Inclán y las artes*, eds. de Santos Zas, Serrano Alonso y De Juan Bolufer (2012).

³ Para la historia de este excepcional Legado, su proceso de transmisión a la USC y la descripción de sus contenidos, ver Santos Zas (2008 y 2012).

el monumental archivo familiar Valle-Inclán/Alsina, que abre nuevas vías a la investigación valleincliniana, entre otras razones, porque significa contar con el primer eslabón de la cadena genética de algunas obras editadas, cuyas galeras corregidas también se conservan, así como –y es su principal novedad– los borradores de obras inéditas, de cuya existencia no se tenía noticia.

De las dos vertientes que conforman nuestro archivo, el actual *corpus* impreso incluye, en un arco de fechas que abarca desde 1888-1936, una serie de materiales que me limito a enumerar, ya que no puedo ni siquiera asomarme a la compleja casuística editorial ni al peculiar sistema de escritura y publicación del escritor (Santos Zas, 2013), con largas prehistorias, rastreables en la prensa y numerosas ediciones y emisiones de una misma obra, que presentan siempre variantes entre sí de mayor o menor calado y que, a los efectos que ahora interesa destacar, conlleva hacer acopio de múltiples versiones. En términos generales contamos con:

1. Más de trescientos textos, correspondientes a ediciones en vida del autor, antologías, prólogos y traducciones de sus obras, que en imágenes digitalizadas alcanzan a día de hoy la cifra de 36.346.
2. A este fondo editorial se agregan las publicaciones periodísticas, es decir, decenas de textos de creación aparecidos en la prensa, sean autónomos o dependientes de obras mayores.
3. De la constante indagación en la prensa proceden igualmente cientos de conferencias y discursos, entrevistas y encuestas y epistolario.
4. Contamos también con un fondo iconográfico, compuesto por fotografías, y caricaturas (que forman parte del archivo personal de Serrano Alonso). Y en el mismo terreno de la imagen, hemos reunido la mayoría de las ilustraciones y motivos ornamentales de las ediciones del autor (ver un amplio muestrario en el citado Portal de la Cátedra Valle-Inclán).
5. El GIVIUS en su indagación hemerográfica ha rastreado la recepción de la obra de Valle-Inclán en su tiempo, con atención particular a sus estrenos teatrales.
6. Por último, hemos reunido numerosas semblanzas y retratos literarios de Valle-Inclán, realizadas por periodistas, intelectuales y amigos.

En suma, estamos hablando de más de cuatro mil quinientos documentos distintos, que traducidos a imágenes digitales sobrepasan las 46.000. De modo que el total de los materiales impresos supera la cifra de 80.000 imágenes.

Desde muy pronto consideramos necesario que el material de prensa, conservado en archivos papel, se fuese incorporando a bases de datos para su posterior consulta y así se hizo ya en los años 90. Y en el caso de las ediciones valleinclinianas, antes incluso de esa fecha –siendo director del Proyecto Luis Iglesias–, utilizamos medios ópticos (Clarity) para la reproducción de imágenes, y OCR (“OmniPage”) para la conversión de dichas imágenes en texto. Este último trabajo se perdió en un intento posterior de trasvasarlo a nuevos sistemas y hubo que repetir el proceso íntegramente. Pues bien, si estos medios fueron hace dos décadas innovadores, es evidente que a estas alturas resultan insuficientes para dar respuesta a nuestras necesidades y objetivos.

Por lo que respecta al Archivo Valle-Inclán/Alsina, contiene lo que el propio escritor calificó de “pruebas de aguafuerte” (Valle-Inclán, 1993 [1924]: 160) ma-

nuscritos de trabajo, materiales en “ebullición” (Grésillon, 1994: 33). Y al decir manuscrito, lo utilizo en sentido genérico y uso el término borradores para señalar la fase redaccional de un autógrafo, un estado de su génesis, sea cual sea éste⁴. Pues bien, en este archivo familiar se conservan no sólo borradores de obras literarias de Valle-Inclán editadas, sino documentos rigurosamente inéditos: vocabularios o listas de palabras, cuadernos de viaje, notas documentales, esquemas, bocetos, resúmenes, ensayos redaccionales más o menos textualizados, versiones de una misma secuencia con tachaduras, sustituciones, adiciones y alteraciones o frecuentes variantes autoriales, copias autógrafas... de poemas, de piezas teatrales, cuentos, novelas, conferencias, pero también numerosas cuentas y facturas, que nos hablan de los detalles de la vida doméstica del escritor o de su relaciones con el mundo editorial... Se trata, en síntesis, del taller del escritor: más de cinco mil páginas autógrafas.

Pero este archivo guarda, además, galeradas corregidas por Don Ramón, ediciones revisadas y anotadas de su puño y letra y algunos mecanoscritos teatrales orientados a la representación escénica. A todo ello hay que añadir numerosos grabados y dibujos originales de artistas contemporáneos destinados al diseño gráfico de sus obras; más de un centenar de cartas originales, que incluyen en ocasiones a emisor y destinatario.

Es ingente la tarea que deparan estos fascinantes materiales⁵, que ayudan a resolver la inmensa laguna, que hasta su hallazgo existía respecto de los procesos de creación de la obra de Valle y sus estrategias de escritura –aproximarse a su *modus componendi*–, que habíamos basado en los pre-textos de muchas de sus obras, publicados en la prensa (con numerosas versiones), y en las ediciones sucesivas, con abundantes variantes textuales, publicadas antes de 1936, sobre las que se han apoyado las ediciones críticas hasta el presente.

⁴ Sobre las definiciones dadas a los términos “manuscrito” y “borrador”, véase Grésillon (1994: 71 y 74). Por su parte, Javier Blasco (2011: 10), siguiendo a Iréne Fenoglio (2007: 8-34), opta por el término *texto* (en cursiva) en oposición a ante-texto, para designar, en el primer caso, el “enunciado o conjunto coherente de enunciados orales o escritos tal como se recogen en la ‘copia final para la imprenta’ o en la edición impresa autorizada”; mientras que el ‘ante-texto’ remite a “las versiones que se contienen en todos los borradores previos a la copia para la imprenta”. Por otra parte utiliza ‘manuscrito’ en sentido genérico, al igual que se usa aquí, para definir el soporte sobre el cual se producen todas aquellas metamorfosis, resultantes del trabajo del escritor (psíquico, gestual y cognitivo), que deja huellas gráficas materializadas sobre el manuscrito. El ante-texto es el espacio propio de trabajo de la crítica genética.

⁵ Afrontar su estudio no ha sido un proceso sencillo y en ocasiones nos hemos visto en la tesitura de tener que rectificar lo ya hecho, pues inicialmente (el primer convenio data de diciembre de 2007) dispusimos del fondo documental exclusivamente en soporte digital, cuyas condiciones no permitían una lectura y transcripción fiables. A finales de noviembre de 2009 se procedió al depósito de los originales del archivo Valle-Inclán/Alsina en la USC y el paso inmediatamente posterior fue acondicionarlos de acuerdo con las normas al uso para la conservación de este tipo de materiales, tarea metódica, que se prolongó hasta finales de 2010, de la que se ocuparon los responsables del fondo antiguo de nuestra Biblioteca Xeral, en cuya caja fuerte se custodian. Contar con financiación fue requisito imprescindible para proceder a su nueva y completa digitalización (marzo-junio 2012), esta vez acudiendo a medios profesionales, que garantizaban su fiel reproducción. Este proceso ha supuesto una reorientación del optimista plan de trabajo inicial, a la que también ha contribuido el paralelo proceso de aprendizaje de los integrantes del GIVIUS, orientado a afrontar la problemática teórico-práctica del estudio y edición de los manuscritos literarios, que en el caso de Valle-Inclán carecía de antecedentes, si exceptuamos la edición de *Mi Bisabuelo*, antes citada, depositada en la USC desde el 21 de junio de 2013 por voluntad de su propietario Jorge Devoto Valle-Inclán.

Si el estudio del archivo del escritor proporciona información muy valiosa, también suscita numerosos problemas: desde su datación, hasta los criterios para priorizar los varios borradores de una misma página, es decir, los diferentes estados redaccionales, cuyo orden de escritura se ignora. Cada borrador –soslayo ahora el concepto de *dossier* genético– exige un tratamiento individualizado que dé respuesta a los específicos problemas que conlleva. De ahí la necesidad de acudir a métodos, que desde focalizaciones atentas a estas realidades documentales y a sus concretos problemas editoriales, no sólo les confieran el valor que tienen para el conocimiento de los procesos de escritura de un autor, sino que proporcionen los recursos metodológicos idóneos para afrontar su carácter inestable, distintos aquellos de las herramientas que se utilizan para el estudio de la obra textualmente fijada.

El GIVIUS desarrolla líneas de investigación consolidadas, articuladas en torno al *corpus* impreso y manuscrito del escritor, que se engarzan en un mismo plan integrador, como diré después, al tiempo que afronta los retos de la aplicación en el campo de las Humanidades de recursos tecnológicos e informáticos *ad hoc*.

En este punto, y al igual que sucede con el *corpus* impreso, se ha realizado la digitalización del Legado Valle-Inclán, ya mencionada, con un resultado cuantitativo de 14.462 imágenes, que incluyen el recto y verso de cada hoja o cuartilla autógrafa de los diversos estados redaccionales que presentan los manuscritos de obras que finalmente dio a la estampa o que han permanecido inéditas; pero también la totalidad de las galeradas con correcciones autógrafas de obras que no llegó a editar; la *editio princeps* de otras con anotaciones manuscritas; el epistolario del escritor vilanovés; grabados y dibujos con destino al diseño de su *Opera Omnia*; y, por último, otros documentos: glosarios, notas de genealogía familiar, anotaciones históricas o esquemas de conferencias.

La digitalización de todo este material era absolutamente imprescindible para abordar, en primer lugar, su indexado y proceder después a su transcripción, estudio y edición. Recordemos que en el ámbito de la filología hispánica moderna y contemporánea, los editores de textos apenas encontraron formas para explorar y valorar mejor la gran riqueza material que contienen los archivos de autor. Los borradores y antetextos se suelen transcribir con las pautas de la crítica textual o ilustran simplemente algunas ediciones críticas, dando fe del paso del filólogo por los archivos del autor.

No es necesario recordar al respecto la labor realizada desde los años setenta por el ITEM (Institut des Textes et Manuscrits Modernes, París), que viene elaborando un protocolo de descripción, transcripción y estudio de los borradores modernos y contemporáneos, más acorde con la naturaleza de un texto “en proceso”. Es decir, los manuscritos de autor requieren dar respuesta a su específica problemática, diferente a la del texto impreso, acudiendo a la genética textual, que proporciona herramientas adecuadas para el estudio de los procesos de creación y sus particulares estrategias.

Nuestro desafío ante este inmenso archivo documental, su estudio y edición requiere no sólo disponer de conocimientos *ad hoc*, sino de medios para su uso y consulta con la mayor eficacia y rentabilidad. Y es aquí donde hemos acudido, sin renunciar a otros, a los recursos tecnológicos e informáticos en su aplicación al campo de las Humanidades, que presupone contar con la colaboración de técnicos en informática, pero también, en el caso de la catalogación, estudio y edición del fondo manuscrito original del escritor, requiere una formación específica de los integrantes del GIVIUS, que desde 2008 sigue siendo una meta,

que tiene su último reflejo en la participación de dos de sus integrantes en el taller "Humanidades digitales: enfoques y aplicaciones" (A Coruña, 8-10 de julio de 2013), pero cuenta con precedentes. El primero se presentó en 2009, recién depositados los manuscritos originales en la USC, con la invitación para participar en el "II Congreso Internacional de Manuscritos de escritores contemporáneos", organizado por CiLengua. Un foro que supuso, por una parte, dar a conocer nuestro Archivo, y, por otra, debatir las propuestas más significativas en este campo con especialistas de varios países. A raíz de este Congreso surgieron otras propuestas del mayor interés para nosotros: en primer lugar, la colaboración con el seminario que impulsa el ITEM "Manuscritos hispánicos XIX-XXI", y dirigen Bénédicte Vauthier y Fatiha Idmhand (enero, 2012); y meses después, la asistencia con dos integrantes del Grupo al "Workshop Digital Text/Genetic Document. Text Encoding Initiative and Genetique (TEI)"⁶, organizado por la prof^a. Vauthier en la Universidad de Berna (septiembre, 2012), cuyo objetivo era la aplicación de los avances tecnológicos e informáticos al estudio y codificación de manuscritos, que se revela fundamental para difundir este patrimonio a gran escala, siendo en el ámbito hispánico una referencia la labor pionera del equipo de la Biblioteca Virtual Miguel de Cervantes, así como el proyecto "Biblioteca Digital Hispánica" de la Biblioteca Nacional de España.

La reunión de Berna fue un fecundo intercambio de experiencias en el terreno teórico, práctico y técnico sobre el manuscrito contemporáneo, que meses después tuvimos ocasión de verificar de la mano de Bénédicte Vauthier, profesora invitada en el máster de "Edición de textos" del Dpto de Literatura española, Teoría da Literatura e Lingüística Xeral, de la Facultade de Filoloxía (enero, 2013), al que asistieron miembros del GIVIUS. Este seminario se centró en el análisis y soluciones dadas a los problemas planteados por el complejo *dossier* genético de *Paisajes después de la batalla*, de Juan Goytisolo (Vauthier, 2012), cuyas pautas de trabajo constituyen un referente para el que nos ocupa⁷.

En nuestro proceso de aprendizaje, todavía abierto, han sido estos pasos esenciales para poder afrontar con responsabilidad y rigor la tarea encomendada por los Herederos de Don Ramón, conducente en último término a la edición de los materiales autógrafos, cometido que requiere su transcripción previa. Ello implica sistematizar una metodología de representación de estos materiales (las tres que se manejan habitualmente son la transcripción diplomática, semi-diplomática y diplomático-interpretativa), aunque este objetivo no excluye la realización de estudios concretos y su difusión, como venimos haciendo desde el momento mismo en que accedimos a este fondo, en revistas y foros internacionales, o su incorporación a las tesis de doctorado nuevas o en proceso. Por otra parte, no es necesario esperar a la transcripción íntegra de los materiales del citado Legado, pues algunos de los textos son unidades redaccionales acabadas y autónomas, que permiten su edición a corto plazo.

⁶ Los investigadores (filólogos e informáticos) integrados en el programa "Text Encoding Initiative" (TEI), que concibieron el sistema de marcado de texto en XML-TEI –sistema que permite la producción de libros digitales–, tratan de adaptar estas herramientas informáticas al objeto de conseguir una transcripción fiel de los manuscritos contemporáneos.

⁷ La prof^a. Vauthier está trabajando en la construcción de un sitio web, en el que se puede seguir –ha tenido la gentileza de mostrárnoslo– el proceso de composición de *Paisajes después de la batalla* a través de un texto en movimiento. En la actualidad participamos en un proyecto conjunto, subvencionado por el FNS suizo.

Esta tarea editorial enlaza con una de las líneas de trabajo del GIVIUS más constantes: la elaboración de ediciones críticas y anotadas de la obra impresa del escritor. Inherente a esta labor es la recopilación de las sucesivas ediciones de cada uno de los textos del autor, al igual que sus numerosas versiones periodísticas, incorporadas, mediante la oportuna reelaboración, a obras mayores o a diversas colecciones, que, a su vez, modifican sus contenidos en sucesivas ediciones, hoy íntegramente digitalizadas.

De ahí también nuestro interés en continuar la indagación en la prensa en busca de todos aquellos textos de creación, aparecidos en periódicos y revistas, que forman parte de la génesis de las obras del escritor; continuar los estudios biográficos, que comporta ampliar la indagación hemerográfica en busca de materiales que los avalen. La abundante prensa periódica de la Edad de Plata y la foránea proporcionan información sobre la biografía de Valle-Inclán, su pensamiento estético y político: entrevistas, conferencias, retratos literarios, cartas y noticias varias, imprescindibles para alcanzar las labores propuestas.

Proseguiremos, de igual modo, el análisis de la recepción crítica contemporánea de la obra del escritor, la acogida de sus estrenos, fundamental para valorar sus éxitos y fracasos, el "impacto" en el panorama literario y cultural contemporáneo y su proyección innovadora.

Nuestro trabajo, que tiene asimismo una significativa vertiente hermenéutica, requiere la continua actualización de la bibliografía secundaria, constantemente incrementada, razón que justifica la preparación y puesta al día de repertorios bibliográficos, que se actualizan en cada uno de los sucesivos volúmenes del *Anuario Valle-Inclán*.

En suma: mientras que el Proyecto subvencionado por el MEC presta atención primordial a los manuscritos del Legado Valle-Inclán y a la coordinación de todas las líneas de investigación que se derivan del *corpus* manuscrito e impreso valleinclaniano; la segunda de las líneas de trabajo del Grupo se viene desarrollando desde 2009 en el Campus de Lugo, a raíz de disponer de un Proyecto de investigación autonómico, subvencionado por la Xunta, dirigido por Serrano Alonso, que a partir de 2012 ha podido continuar su plan de trabajo en el marco del Plan de Consolidación de Grupos de investigación, de la Xunta de Galicia.

En nuestro ambicioso proyecto de trabajo ha sido objetivo prioritario y cumplido la completa digitalización de nuestro fondo documental y bibliográfico, como ya he dicho; ahora dos metas ocupan nuestro esfuerzo: continuar el indexado de todos los materiales valleinclanianos –manuscritos e impresos– recopilados a lo largo de los últimos veinte años; y dar cima a un Catalogador informático, que permitirá no sólo contener toda la información acerca de los documentos reunidos, sino ofrecer las citadas digitalizaciones y facilitar búsquedas integrales de todo tipo de información. Además, este Catalogador permitirá su utilización en red de uso restringido al GIVIUS, sin que descartemos en un futuro no lejano el acceso público a estos recursos.

Se ocupan de explicar a continuación este plan de trabajo, Francisca Martínez, doctoranda contratada del Proyecto, en "El archivo manuscrito digital valleinclaniano"; y Rosario Mascato, titular de un contrato posdoctoral, en "Gestión y tratamiento del *corpus* impreso en el Archivo Digital Valle-Inclán". Ambas han estado en el origen y desarrollo de los respectivos archivos y darán cuenta de las herramientas informáticas de digitalización, catalogación, gestión y consulta de los materiales literarios, que responden al proceso de escritura y publica-

ción de la obra de Valle-Inclán, así como a su recepción crítica, sin soslayar la biografía del autor.

2. EL ARCHIVO MANUSCRITO DIGITAL VALLEINCLANIANO

Desde su creación, como se ha dicho, el GIVIUS ha ido haciendo acopio de lo que hoy es un voluminoso fondo bibliográfico y documental –impreso y manuscrito– que ya en la década de los 90 hizo necesaria la puesta en marcha de herramientas digitales, en concreto la primera base de datos en *access* para indexar entrevistas, cartas y conferencias valleinclinianas, creada, según las pautas de la dirección del GIVIUS, con la colaboración del profesor Gómez Segade, que ha sido útil hasta la puesta en marcha del archivo que ahora presentamos.

Por lo que se refiere al fondo manuscrito, hasta el año 2008 disponíamos de copias digitales de documentos autógrafos, raramente de creación (cartas, informes, memorias, dedicatorias, algunos poemas y un cuento...), que se vieron extraordinariamente incrementados, tanto en la vertiente creativa como en la documental, gracias al Legado Valle-Inclán Alsina (Fig. 1), cuyos originales están hoy depositado en la USC, según lo antes expuesto. A este archivo se han ido sumando otros materiales manuscritos adquiridos por el GIVIUS o donados a

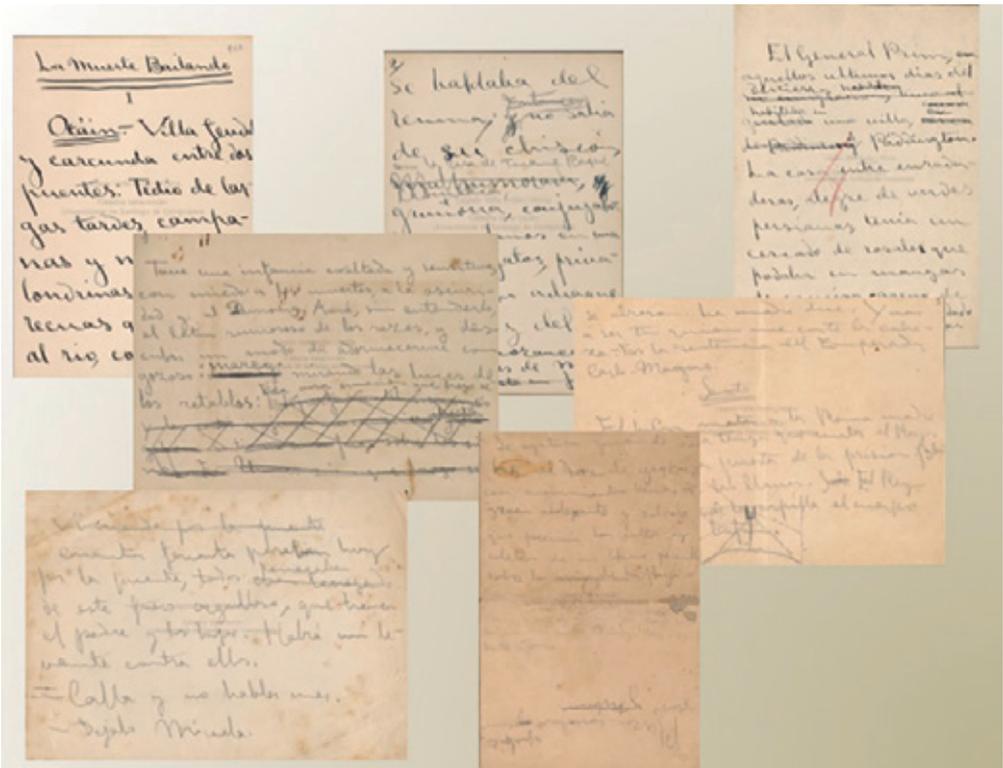


Fig. 1. Selección de páginas sueltas de varios borradores inéditos del Legado Valle-Inclán Alsina/Cátedra Valle-Inclán de la USC.

este por sus propietarios, de manera que para poder estudiarlos y editarlos se imponía la creación del Archivo Digital de Manuscritos Valleinclanianos, herramienta imprescindible para ofrecer el indexado de estos materiales en una base de datos relacional, que permitiese, en primer lugar, la ordenación, descripción y consulta de ese excepcional legado, para proceder, una vez culminada esta fase, a su integración en el Catalogador que se describe en el tercer apartado de este trabajo.

A tal fin, han sido determinantes en lo que se refiere a nuestro aprendizaje, los foros académicos arriba mencionados, en los que he participado con otros integrantes del GIVIUS, de los que destacaría por su carácter eminentemente práctico el *Workshop Digital Text/Genetic Document. Text Encoding Initiative and Genetic Edition*, celebrado en Berna (2012), bajo la coordinación de la profesora Bénédicte Vauthier. En esa ocasión las sesiones expositivas se combinaron con intensas sesiones prácticas sobre la aplicación de TEI, dirigidas, en nuestro grupo, por la profesora Elena Pierazzo, del Department of Digital Humanities del King's College de Londres, que generosamente nos facilitó numerosos materiales de trabajo y expuso la compleja praxis de codificación de manuscritos, ofreciéndonos una amplísima casuística de procesos de marcado TEI sobre textos de tipología genérica diversa. A esta experiencia se ha sumado ahora el workshop "Humanidades Digitales: enfoques y aplicaciones" (Universidade da Coruña, 8-10 de julio), orientado sobre todo al uso de herramientas digitales en el ámbito de las humanidades.

De cada uno de esos encuentros hemos extraído fórmulas y prácticas útiles para el tratamiento digital del fondo manuscrito de Valle-Inclán, a la vista de otras experiencias eficaces a la hora de manejar archivos similares en bases de datos que no sean meros repositorios, sino que sirvan de manera eficaz y rentable a la investigación del Grupo; es decir, que nos permitan indexar los materiales, formular campos con criterios genéticos y hacer uso de los metadatos para su consulta y posterior edición.

Para crear nuestra base, en este caso sin apoyo técnico externo, el primer paso era la indexación de materiales, y para ello hemos tomado como punto de partida el inventario de los documentos originales que conforman el Legado Valle-Inclán Alsina, elaborado por Joaquín del Valle-Inclán y revisado posteriormente con la profesora Santos Zas.

Este inventario consigna los materiales tal como fueron organizados por su propietario –Carlos del Valle-Inclán Blanco– y han llegado a nuestras manos, de modo que desconocemos su estado en origen. De hecho, el hijo del escritor realizó un primer intento de ordenar la totalidad de los borradores autógrafos con criterios temáticos (según se desprende de nuestra propia confrontación), ya que no están datados. De manera que los agrupó en carpetas comunes, numeradas y en su mayoría designadas con títulos –no siempre autoriales–, que acogen, a su vez, una serie variable de subcarpetas, que se enumeran en dependencia de la principal, cada una de las cuales mantiene cierta unidad de sentido, pero no con las que le preceden o le siguen. Quiero decir, que faltan engarces entre unas y otras, lo cual no siempre ni necesariamente significa que se hayan perdido, sino que en ocasiones se han ubicado en otras carpetas de contenido dispar, posiblemente como resultado de una lectura incorrecta o sencillamente porque, al tratarse de páginas sueltas en su mayoría, las posibilidades de que se altere el orden dentro y/o entre las subcarpetas, que integran cada carpeta, se multiplican.

Esta organización se ha respetado escrupulosamente al encargar a una empresa profesional la digitalización de la totalidad del Legado Valle-Inclán, proceso supervisado por la responsable del GIVIUS, que concluyó en junio de 2012. Esta nueva digitalización en verso y recto, que alcanza un alto grado de fiabilidad, constituye el segundo pilar para la construcción de la base de datos.

La creación de este archivo digital, que en último término deberá incorporarse al Catalogador "Archivo Digital del Grupo de Investigación Valle-Inclán", para que alcance su pleno sentido y rendimiento, responde a los siguientes objetivos:

1. Facilitar los materiales indexados de acuerdo con la organización original, pero subsanando sus errores en la ordenación de contenidos.
2. Ofrecer una descripción material y textual de cada una de las unidades que conforman cada carpeta, creando campos *ad hoc* (Fig. 2), etapa ineludible en el proceso de estudio y edición de los manuscritos.
3. Proporcionar, en una segunda fase, la transcripción diplomática del documento, que deberá quedar respaldada por un archivo txt, necesario para la posterior incorporación al citado Catalogador.

**Legado Manuscrito
Valle-Inclán**

Sinatura:

Denominación prooia:

Descripción:

Formato:

Soporte:

Instrumento:

Recto Verso

Comienzo de página:

Final de Página:

Transcripción:

Texto en TEI:

Fecha de creación:

Revisión:

Fig. 2. Registro de FileMaker/Biblioteca de documentos, diseñado para el Archivo digital valleinclaniano.

4. Complementariamente, crear una versión del texto, marcado con *Oxygen editor* (Fig. 3), en TEI (*Text Encoding Initiative*), consorcio que desarrolla y mantiene un estándar para la representación de textos en formato digital y ofrece un conjunto de directrices que especifican métodos de codificación de textos de lectura mecánica, sobre todo en las humanidades, las ciencias sociales y la lingüística. Desde 1994, las Normas TEI han sido utilizadas, además de por bibliotecas y museos, por editores e investigadores, pues aunque la adquisición de los conocimientos necesarios para su uso con pleno rendimiento no está exenta de complejidad, también creemos que las posibilidades que ofrece son enormemente rentables en función de compartir y trabajar en línea con una versión xml del documento (Fig. 4), que introduciría metadatos como parte de su información, viabilizando un sistema de búsquedas eficaz, y en consecuencia la elaboración de un *dossier* genético.

La codificación TEI permitirá aplicar estructuras genéricas para cada documento a través de la herramienta *ROMA* (Fig. 5), generador de esquemas determinados, con definiciones de elementos y atributos para cada tipología textual, en coherencia con la diversidad de materiales contenidos en el Legado Valle-Inclán Alsina.

En suma, tras el indexado, se obtiene un archivo por cada carpeta, que incluye tantos registros como imágenes estén contenidas en ella. Estos son fmp (FileMaker pro), si deseamos trabajar directamente en la base de datos, o pdf/Excel, sobre los que se podrán realizar búsquedas sin riesgo de modificar el contenido de los campos de la base.

Con el fin de agilizar este proceso, hemos optado por una base de datos con gestor integrado, creada a través de la aplicación FileMaker, una base de datos

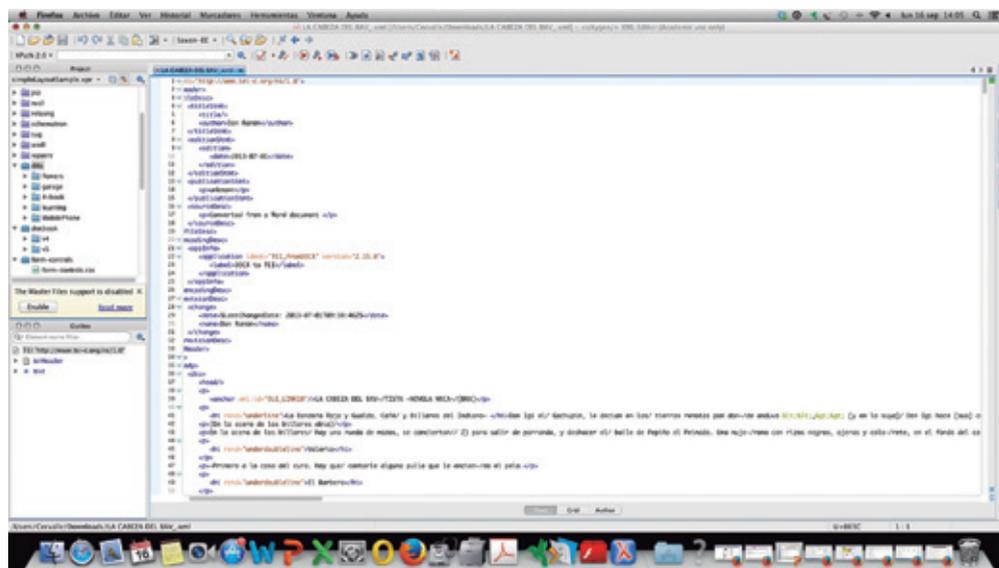


Fig. 3. Captura de pantalla de texto marcado con *Text Encoding Initiative* (TEI), con el programa *Oxygen Editor*, modalidad de visualización "Text".

relacional que permite crear formularios, informes, etc., muy similar al modelo de Microsoft Access, en la que elegimos una solución prediseñada –Biblioteca de documentos–, modificando íntegramente los campos existentes para adecuarlos a las necesidades antes descritas.

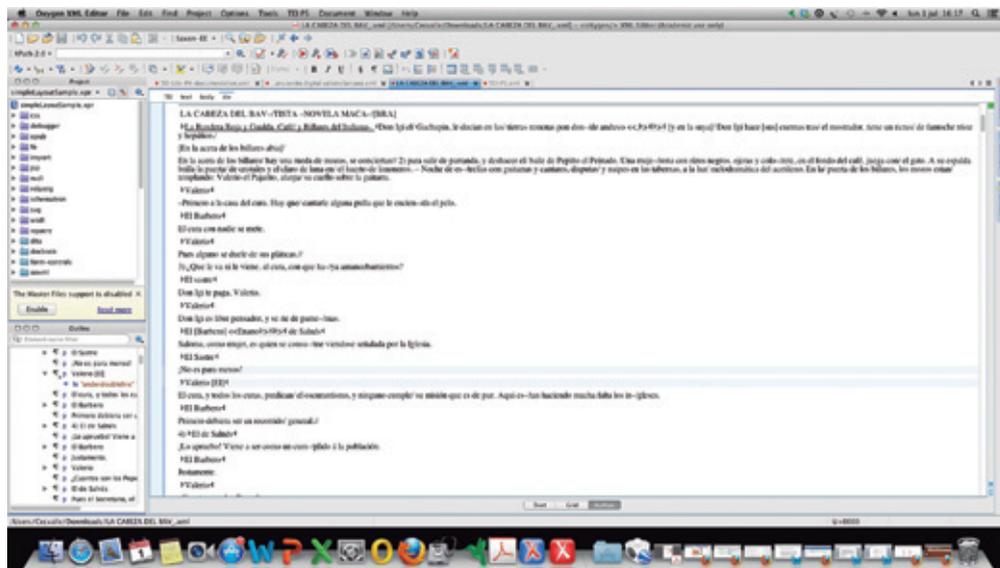


Fig. 4. Captura de pantalla de texto marcado con Text Encoding Initiative (TEI), con el programa Oxigen Editor, modalidad de visualización "Author".

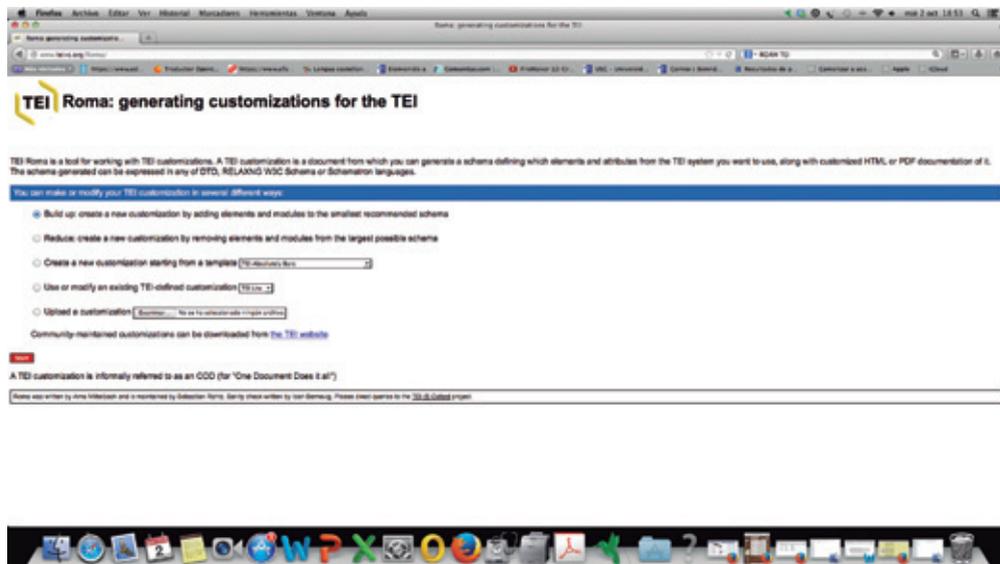


Fig. 5. Captura de pantalla de ROMA, página de inicio de generador de diseños para marcado TEI.

FileMaker puede importar datos de excel, archivos xml o fuentes ODBC y los informes generados pueden ser exportados en excel o PDF. En una fase más avanzada de nuestro trabajo utilizaremos Aperture, aplicación profesional de edición de imágenes digitales, que permite exportar los metadatos de las imágenes en un archivo de texto.

La versión servidor de la aplicación permitirá incorporar en futuro las bases de datos en el mencionado Catalogador. Hasta entonces, con la versión básica, tenemos la posibilidad de compartir las bases de datos con otros usuarios en una red local.

A la hora de elaborar los campos que figuran en cada uno de los registros, hemos acudido a criterios genéticos, que nos permitan una descripción material del documento. Así, contemplamos en nuestra base los siguientes campos:

1. **Signatura:** designación existente en el archivo original, que identifica cada carpeta e imagen en nuestro index.
2. **Denominación propia:** permite enmendar posibles denominaciones erróneas del punto anterior.
3. **Descripción:** en el registro inicial de cada archivo figura la adscripción genérica del documento: teatro o texto dialogado, narrativa, poesía, ensayo, glosarios, notas genealógicas, históricas, literarias...; se indica también el tipo de documento (borrador autógrafo, esquema, boceto, galerada...).
4. **Formato:** medidas del soporte establecidas con un medidor angular de editor de imagen.
5. **Soporte:** hoja suelta, cuartilla, cuaderno, bloc... Se consignan problemas que afectan al estado del soporte.
6. **Instrumento:** pluma o lápiz (negro, rojo, azul...).
7. **Espacio gráfico:** se indica su uso en recto y/o verso y recto.

En todo momento se puede visualizar la imagen que corresponde al elemento descrito, con posibilidad de exportarla fuera de la base.

En lo que al contenido del propio documento se refiere, sea cual sea su tipología, categoría y género, hemos diseñado cuatro campos:

1. **Palabras iniciales** de cada unidad digital, que ya figuraban en el inventario de los manuscritos originales.
2. **Palabras finales** de cada unidad digital, inexistentes en el citado inventario.

Ambos están destinados a la detección de incoherencias en la secuencialización del archivo.

1. **Transcripción**, según pautas descritas en el apartado 2 de los objetivos del archivo digital.
2. **Texto en TEI**, véase, asimismo, la explicación en el apartado 3 de los objetivos del archivo.

Estos dos últimos campos deben posibilitar al investigador una base textual fiable, en tanto uno y otro sistema están concebidos como punto de partida esencial para cualquiera de los estudios y modalidades editoriales que afrontemos en futuro.

El diseño por el que hemos optado, siempre perfectible, es sencillo pero eficaz, con la intención de agilizar su implementación en coherencia funcional con el modelo relacional del Catalogador y así tratar de lograr la creación de una macrobase de datos valleinclaniana, cuya aspiración es ofrecerse, en el contexto de las humanidades digitales y de la transferencia del conocimiento, como un modelo para la organización digital de los archivos de autor.

3. GESTIÓN Y TRATAMIENTO DEL *CORPUS* IMPRESO EN EL ARCHIVO DIGITAL VALLE-INCLÁN

Uno de los principales objetivos del Grupo de Investigación Valle-Inclán a lo largo de su historia ha sido ejercer una exhaustiva labor de recuperación de *documentación primaria* que, más allá, de la propia obra literaria del escritor, nos permitiese reconstruir su trayectoria en el marco del campo literario español de su época. Así, en los diversos estudios y trabajos publicados por los miembros del equipo, estos materiales se constituyen como eje para diversos abordajes metodológicos que, en unas ocasiones priman el análisis del proceso de elaboración literaria del autor (el denominado *taller del escritor* siempre considerado como *work in progress*); en otras, el estudio de las condiciones editoriales, institucionales e ideológicas que afectan a su producción en el marco del campo literario de inicios del siglo XX, con el afán de determinar el valor más o menos central o periférico de sus propuestas estéticas, identitarias o políticas con respecto a los repertorios en vigor en dicho campo. Independientemente de uno u otro enfoque, estos trabajos pretenden –en su conjunto– clarificar los límites existentes entre las verdaderas dimensiones de Valle-Inclán en la España de su tiempo, frente al conocimiento construido posteriormente por sus exégetas en el campo de la crítica valleinclaniana.

Para llevar a cabo esta tarea, es necesario trabajar con una extensa tipología de materiales que amplían la perspectiva generada por la obra literaria, en primera instancia, y ofrecen datos primarios en torno al desempeño de la labor del escritor: versiones de obras (previas o posteriores a la edición en libro); textos sobre estética y poética (bien en entrevistas, bien en artículos, conferencias o discursos); declaraciones públicas sobre asuntos políticos, culturales o sociales (como manifiestos o cartas públicas) a lo que se suman, por supuesto, primeras (y posteriores) ediciones de sus obras, al igual que antologías y colectáneas que recogen sus textos. Sin olvidar la propia recepción de época: reseñas críticas, artículos, materiales sobre sus estrenos teatrales e incluso parodias (por ejemplo, de sus poemas). A todo lo cual se suma un ingente material iconográfico que va desde centenares de caricaturas y una considerable colección de retratos y fotografías, a las ilustraciones que acompañan a sus obras (tanto en formato libro como en las páginas de la prensa). Y, claro está, otros materiales de primera mano, como epistolario, autógrafos o documentos personales.

En consecuencia, nuestro *corpus* de estudio se caracteriza por su variedad, extensión y riqueza en matices (Fig. 6). Se trata, en conjunto, de más de 4.500 documentos, de los que –una vez descontados los textos editados en libro, pequeños autógrafos y epistolario manuscrito– más del ochenta por ciento han sido recogidos de la prensa de la época (gallega, española, europea y americana), conformando así lo que hemos denominado como *corpus mediático* del escritor (Mascato, 2013).

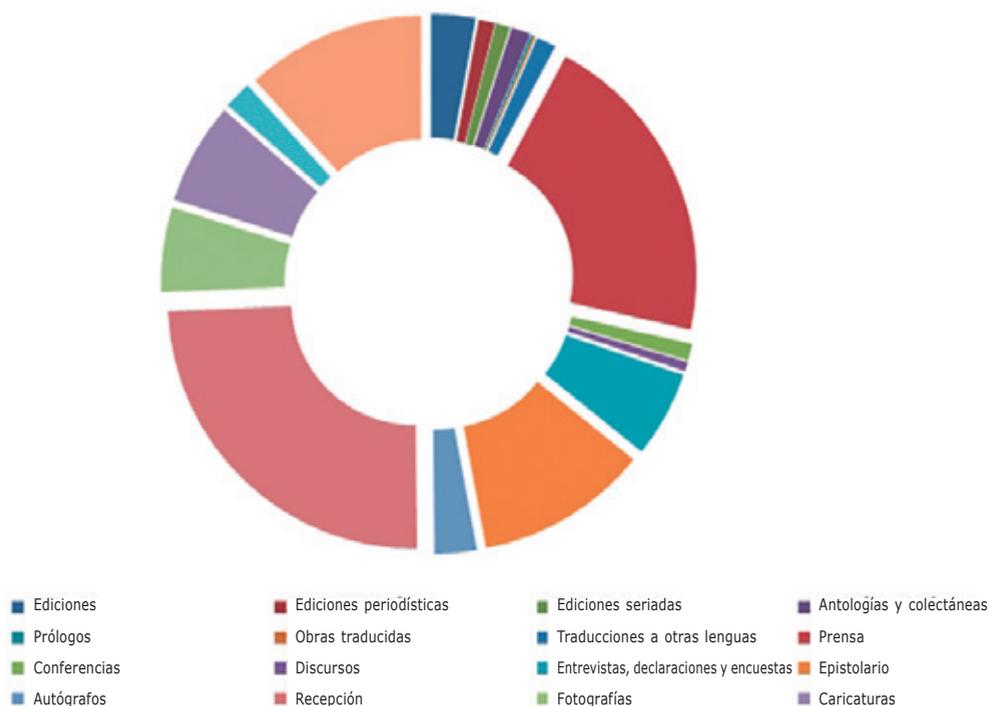


Fig. 6. Corpus impreso completo del Archivo Digital Valle-Inclán.

Este vasto *corpus* impreso requería, dada su complejidad y el estado precario en que se encuentra en muchos casos (libros con encuadernación deficiente, fotocopias con un tamaño o resolución inadecuados, microfilms de difícil acceso...), ser sometido a un proceso de ordenación y catalogación, para su posterior digitalización, una tarea que, con precedentes en el propio GIVIUS que se remontan a los años 90, alcanza su impulso definitivo en 2009 en el marco del proyecto “Valle-Inclán: a prensa e o sistema editorial” (2009-2012), dirigido por el profesor Javier Serrano Alonso, en la Facultad de Humanidades de Lugo, gracias a la financiación del Programa INCITE, de la Consellería de Industria de la Xunta de Galicia.

El proceso de digitalización de todos estos materiales, finalizado hace un año, consolidó las pautas de los procedimientos para el escaneado de textos (de acuerdo con los estándares actuales) –con un escáner adquirido *ad hoc* para el trabajo específico con libros (Plustek OpticBook 3600)–, la preparación de las imágenes individuales de cada página para obtener los mejores resultados de lectura –mediante *software* de edición de imagen (Adobe Photoshop CS3)–, la preparación de los PDF de cada obra o documento en su conjunto (vía Adobe Acrobat 9 Pro Extended) y el posterior tratamiento de cada documento con *software* de reconocimiento de texto (Abby Fine Reader 10 Corporate Edition), para la consecución de PDFs con función de búsqueda textual y señalización por marcadores, así como de documentos TXT pensados como base de gestión y búsqueda para diversas aplicaciones digitales.

En paralelo, a finales del año 2011 iniciamos el proceso de diseño conceptual y técnico de una macrobase de datos cuyos objetivos fundamentales eran:

1. Catalogar la documentación e información de ella derivada.
2. Indexar los documentos.
3. Disponibilizarlos para su consulta *online* y eventual descarga, de momento en un servidor con acceso restringido (al menos hasta 2016).
4. Habilitar la posibilidad de realizar búsquedas (por campo, categoría o incluso textuales).
5. Añadir nuevas tipologías de documentación.
6. Poder analizar esta información en clave textual, relacional, cuantitativa y gráfica en un futuro, mediante la utilización de nuevas aplicaciones y programas.

Ambos procesos confluyeron finalmente en la creación del **Archivo Digital del Grupo de Investigación Valle-Inclán**: una base de datos MySQL, exportable (por tanto) en lenguaje XML, y asimilable en la mayor parte de campos utilizados a los estándares internacionales (DublinCore, Marc21...) propios de bibliotecas y repositorios digitales de consulta pública⁸ (Fig. 7).

Se trata de un macrocatalogador, en la actualidad dotado con diversas tipologías de usuario (superadministrador, administrador de unidad organizativa, catalogador, público), lo que permite simultanear el proceso de desarrollo y *upload* de documentos, con la gestión y tratamiento de los materiales, así como con su consulta.

En su estructura interna, el catalogador incluye diversas unidades organizativas, diseñadas en función de las características de las distintas secciones del *corpus*⁹: antologías y colectáneas, autógrafos, caricaturas, conferencias, dibujos, discursos, ediciones, ediciones periodísticas, ediciones seriadas, entrevistas, epistolario, fotografías, obras traducidas, prensa, prólogos y traducciones. Estas unidades funcionan como tablas relacionadas entre sí, que en algunos casos llevan asociadas otras con información que puede ser individualizada (y consultada) de manera transversal como parte de los intereses investigadores del equipo (lugares de publicación, periódicos y revistas, editoriales, imprentas, instituciones o agentes con que se relaciona Valle-Inclán, por ejemplo).

Cada registro incorpora los datos fundamentales del documento para su localización, así como una imagen del mismo, el archivo PDF correspondiente habilitado para búsquedas y el TXT sobre el que el propio catalogador genera la información textual. Todos estos datos están disponibles para el usuario final del Catalogador (Fig. 8).

Inicialmente ubicado en un servidor en pruebas, una de las incidencias de mayor relieve con que se encontró este proyecto fue el hecho de que la Universidad de Santiago carece, en estos momentos, de espacio en su servidor para

⁸ En la actualidad, el proyecto se encuentra precisamente en fase de adaptación a estos estándares internacionales, llevando a cabo un proceso de revisión del catalogador.

⁹ A estos efectos, hemos seguido, fundamentalmente, las pautas indicadas por Javier Serrano Alonso para su nueva versión de la *Bibliografía de Ramón del Valle-Inclán* –pendiente de edición–, que supone una ampliación y reformulación del *corpus* valleinclaniano en cuanto a la tipología y número de documentos en ella recogidos, frente a lo propuesto en Serrano Alonso y Juan Bolufer (1995).

The screenshot shows the public search interface of the Digital Archive of Valle-Inclán. At the top, there is a header with the USC logo (Universidad de Santiago de Compostela), the title "ARCHIVO DIGITAL Grupo de Investigación Valle-Inclán", and a signature. Below the header, there is a main menu with "Inicio" and "Catalogar". The search area includes filters for "Unidad Organizativa:", "Lugar de Publicación:", and "Año:". There are three dropdown lists: the first contains categories like "Antologías y Colecciones", "Autógrafos", "Caricaturas", "Conferencias", "Dibujos", "Discursos", and "Ediciones"; the second contains locations like "A Pobra do Caramiñal (A Cersúa)", "Alicante", "Amsterdám", "Asunción (Paraguay)", "Bagnollet (Francia)", and "Barcelona"; the third contains years from 1866 to 1872. Below these are search fields for "Fecha: Desde:" and "Hasta:", "Búsqueda por campo:" (with a note "seleccionar palabras por comas"), "Búsqueda en el PDF:" (with a note "seleccionar palabras por comas"), and "Serie:" with two dropdown menus for "Nivel 1:" and "Nivel 2:". A "Buscar" button is located at the bottom right.

Fig. 7. Captura de pantalla del buscador público del Archivo Digital Valle-Inclán.

alojar una macrobase de datos de estas características y sus contenidos (que exceden en mucho los 25 Mb habitualmente disponibles a estos efectos). Esta situación, que no había sido contemplada en el *planning* inicial del proyecto, nos obligó a adquirir un servidor propio, al que la base fue trasladada –tras el período de construcción– en diciembre de 2012.

Desde esta fecha, el Archivo se encuentra en proceso de refinado técnico y de contenidos, en una exhaustiva revisión (en la que están implicados todos los miembros y colaboradores del Grupo), con el objeto de solventar cualquier incidencia que afecte a la efectividad y corrección en los procesos de catalogación y búsqueda: que no se produzcan fallos ni problemas de apertura de archivos, que el constante incremento del volumen de información no tenga incidencias sobre la velocidad y fiabilidad de las consultas, que la información sea indexada y jerarquizada de la manera adecuada con respecto a las prioridades del usuario en la consulta, etcétera.

Sin embargo, conviene indicar que un proyecto de esta entidad no está exento de problemas: en este caso precisamente derivados del trabajo con prensa histórica (grueso de nuestro *corpus*), cuya mayor dificultad reside en el proceso para la obtención del OCR (y, por tanto, de los archivos TXT, eje del trabajo de análisis y edición).

Si bien para el caso de aquellos materiales impresos en formato de libro, coleccionable o revista apenas hemos registrado dificultades de lectura, con esca-

Archivo Digital Valle-Inclán

Poemas de las Rosas. Rosa gitana [Rosa de Oriente]. Rosa de ultramar [Alegoría]



General

Título	Poemas de las Rosas. Rosa gitana [Rosa de Oriente]. Rosa de ultramar [Alegoría]
Publicación	El Imparcial
Lugar (de publicación)	Madrid
Año	1912
Fecha	1912-06-24
Tomo, Nº, Páginas	p. 3
Género	Poesía
Serie	• Claves Líricas / El Pasajero
Ilustración	
PDF	Ver PDF

Fig. 8. Resultado de una búsqueda en el Archivo Digital Valle-Inclán.

Las intervenciones en la corrección del texto, en el de la prensa diaria, concurren toda una serie de elementos que entorpecen la progresión del proyecto y nos han obligado a reconsiderar las prioridades y tareas inmediatas del equipo: el estado deficitario de la mayor parte de los periódicos y revistas originales, en lo referente a la calidad de impresión y el papel utilizado en la época; las dificultades de obtener una digitalización de determinadas cabeceras de prensa, en función muchas veces de su origen o de las instituciones que son depositarias de las mismas (es el caso de las bibliotecas cubanas, por ejemplo, donde la ma-

oría de nuestros materiales han sido obtenidos en fotocopia); o no disponer de una digitalización adecuada, al haber sido conseguidas a través de repositorios digitales (el caso de la Hemeroteca Digital de la BNE) o de la solicitud de escaneados de uno u otro documento por parte de determinados archivos o hemerotecas, que cuentan con una política, medios técnicos y parámetros propios para este tipo de procesos.

Todo esto implica que, a pesar de utilizar las últimas versiones de software de reconocimiento de texto disponibles en el mercado (el mencionado Abby Fine Reader 10), los resultados estén lejos de ser satisfactorios, como se puede observar en los dos ejemplos aportados (Figs. 9 y 10).

Sin embargo, es esta una problemática común a otros proyectos de digitalización de prensa de los siglos XIX y XX. Incluso en aquellos casos en que los equipos de trabajo implicados cuentan con una capacidad de financiación mucho mayor a aquella de que puede disponer un grupo de investigación de las características del GIVIUS.

Un ejemplo reciente de ello lo ofrecía, en 2009, Rose Holley, directora del programa de digitalización de prensa histórica de la National Library of Australia:

In a sample of 45 pages to be representative of the libraries digitised newspaper collection 1803-1954, we found that raw OCR accuracy varied from 71% to 98.02%

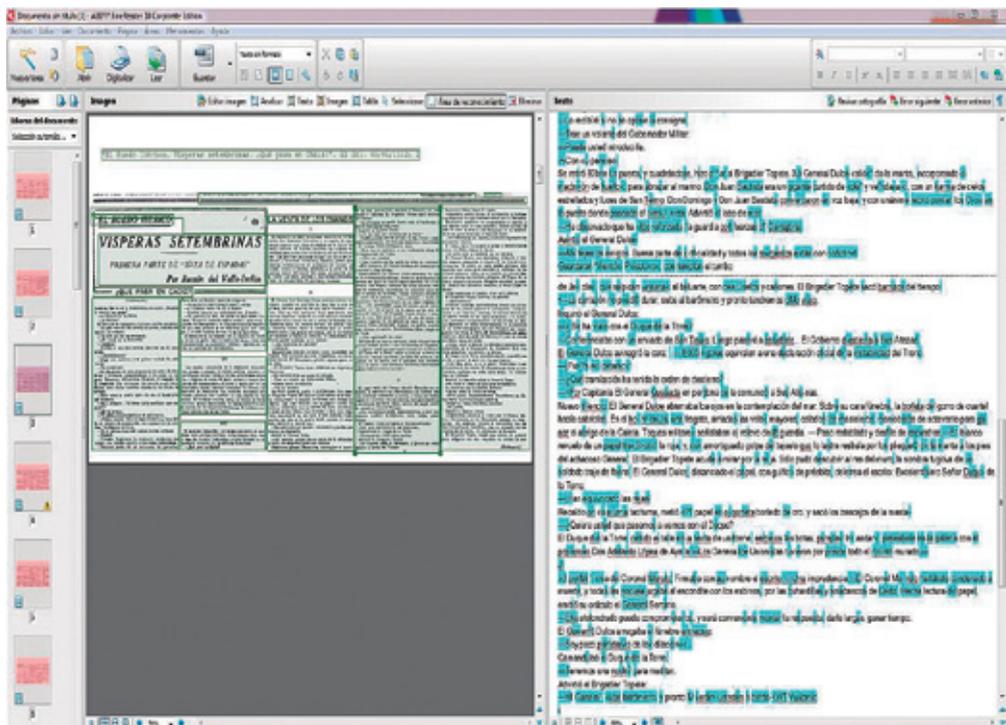


Fig. 9. Captura de pantalla de un proceso de reconocimiento de texto realizado con Abby Fine Reader 10 – Corporate Edition, sobre un fragmento de "Visperas Septentrinas", folleto perteneciente a la serie de El Ruedo Ibérico y publicado en El Sol, de Madrid (1932).

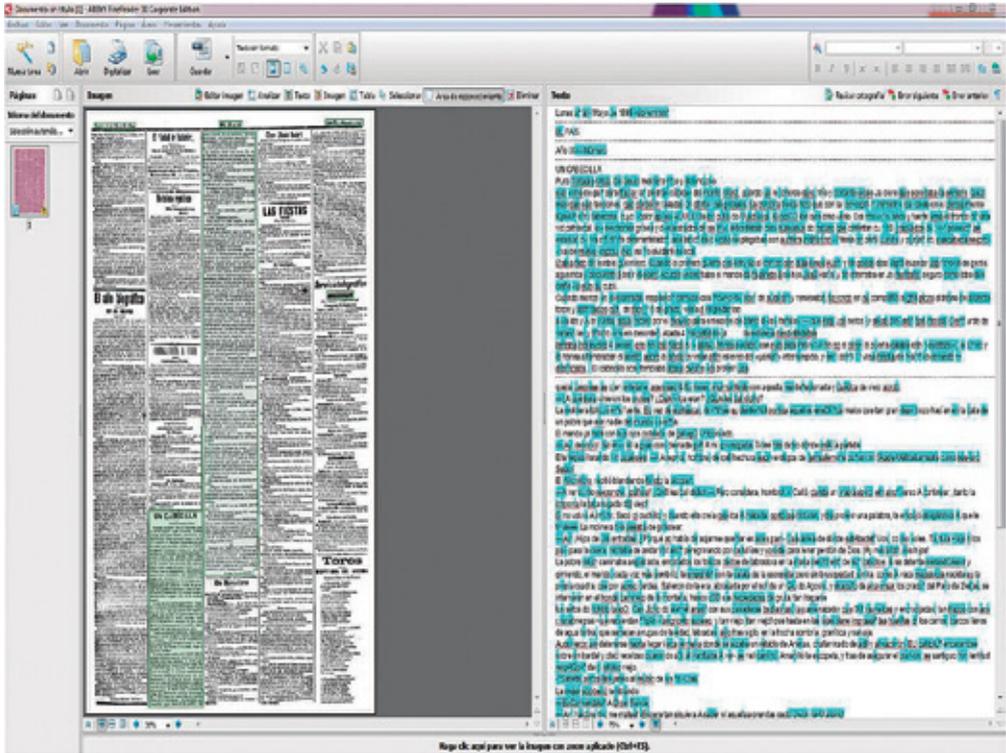


Fig. 10. Captura de pantalla de un proceso de reconocimiento de texto realizado con ABBY Fine Reader 10 – Corporate Edition, sobre el cuento “Un Cabecilla”, publicado en El País (1895).

(character confidence, no correction). At the 71% accuracy rate there would be 145 incorrect characters in an average paragraph of 500 characters. Looking at this the other way round means that 29% of the paragraph would be incorrect. This makes reading and accurate text retrieval difficult, even more so if the incorrect characters are not whole words but characters within many words. The question of what is acceptable has not been answered, but in speaking to other libraries and OCR contractors, it was generally agreed for historic newspapers that when we talk about good, average and bad OCR we mean:

Good OCR accuracy	= 98-99% accurate	(1-2% of OCR incorrect)
Average OCR accuracy	= 90-98% accurate	(2-10% of OCR incorrect)
Poor OCR accuracy	= below 90% accurate	(more than 10% of OCR incorrect)

Este es, por tanto, el gran problema al que se enfrenta un proyecto de nuestras características, para el que en la actualidad, las soluciones pasan no sólo por una mera cuestión de mejora de las condiciones técnicas de los materiales y el software utilizado para la obtención del OCR, sino –sobre todo– por un proceso de colaboración con el usuario del repositorio. Esto es, el “manual

method of humans correcting the mistakes of a machine”, labor a la que nos hemos visto obligados a dedicar parte de los esfuerzos del GIVIUS a fin de conseguir el mayor grado deseable de corrección en los resultados obtenidos (muy especialmente en lo referente a las búsquedas textuales).

A pesar de estas dificultades, son varias las líneas de trabajo que se desprenden de la implementación de esta importante herramienta en el desarrollo de los objetivos del grupo de investigación:

Por una parte, a corto plazo, podremos examinar las condiciones de producción de la obra literaria en el campo literario, enfocando la trayectoria de Valle-Inclán en el mismo (su función y posición, de acuerdo con los parámetros establecidos por Bourdieu, 1992), a través de su relación con agentes e instituciones del mismo campo o de otros campos culturales y políticos. En este sentido, el análisis cuantitativo y relacional de los datos ya contenidos en nuestro catalogador nos permitirán, entre otras cosas, mapear la red de relaciones del escritor con su entorno, siguiendo el modelo de estudios anteriores referidos a casos literarios, como los de Samartin (2010) o Long y So (2013).

Por otra, tendremos la capacidad de profundizar, a medio plazo, en el estudio de la evolución estética y poética del literato a partir de datos empíricos y objetivos, obtenidos de secuencias textuales y concordancias a través de la aplicación del marcado semántico (TEI), lo que establecerá igualmente nuevos parámetros para la elaboración de ediciones críticas y genéticas hipertextuales. Para ello, resulta objetivo esencial del equipo obtener la transcripción (por medios técnicos y humanos) de la totalidad del *corpus* (manuscrito e impreso). El resultado, no menos ambicioso (por complejo) que su fase previa, nos permitirá visualizar en pantalla los diferentes estados creativos de una obra y la evolución, paso a paso, de Valle-Inclán, como productor de discursos e ideas.

En uno y otro caso, la tarea se antoja laboriosa, porque, entre otras cosas, pretendemos contribuir a un cambio de paradigma en el marco de la crítica (valleinclaniana), haciendo uso de esta nueva herramienta digital que es el Archivo Valle-Inclán para la mejor comprensión del hecho literario.

BIBLIOGRAFÍA

- Blasco Pascual, Javier, *Poética de la escritura. El taller del poeta. Ensayo de crítica genética* (Juan Ramón Jiménez, Francisco Pino y Claudio Rodríguez), Valladolid/New York, Universidad de Valladolid/Junta de Castilla y León, Col. Cátedra Delibes, 2011.
- Bourdieu, Pierre, *Les Règles de l'art*, Paris, Éditions du Seuil, 1992.
- Burnard, Lou y Bauman, Syd, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, Charlottesville (Virginia), Text Encoding Initiative Consortium, 2013.
- Fenoglio, Irène, "Du texte avant le texte. Formes génétiques et marques énonciatives des pré-vision textuales", en *Langue française*, nº 155 (2007), pp. 8-34, <<http://www.cairn.info/revue-langue-francaise-2007-3-page-8.htm>>.
- Grésillon, Almut, *Éléments de critique génétique. Lire les manuscrits modernes*, Paris, PUF, 1994.
- Grésillon, Almut, *La mise en oeuvre. Itinéraires génétiques*, Paris, CNRS Editions, 2008.
- Hay, Louis et al., *Les manuscrits des écrivains*, Paris, CNRS, 1993.
- Koenig, Glenn, "The Origin of FileMaker Pro", <<http://www.dancing-data.com/filemakerhist.html>>.

- Holley, Rose, "How Good Can It Get? Analysing and Improving OCR Accuracy in Large Historic Newspaper Digitisation Programs", *D-Lib Magazine*, 15, 3/4 (March/April, 2009), <<http://www.dlib.org/dlib/march09/holley/03holley.html>>.
- Long, Hoyt y So, Richard, "Network Analysis and the Sociology of Modernism", *Boundary 2. An International Journal of Literature and Culture*, 40 (2) (2013), pp. 147-182.
- Marcos, M^a del Carmen, "Los archivos en la era digital", <<http://goo.gl/7tw8Or>>.
- Mascato Rey, Rosario, *Valle-Inclán, poeta moderno no canonizado*, Universidade da Coruña, Servicio de Publicacións, 2013.
- Rodríguez, Javier, "FileMaker Pro 12: base de datos relacional", *Macworld España: la publicación para los usuarios del Mac, el iPhone y el iPod*, nº 224 (2012), pp. 94-95, <<http://www.idg.es/macworld/>>.
- Samartin, Roberto López-Iglesias, "Défices projetivos e estratégias de planificación cultural no campo editorial dum sistema periférico (Galiza: 1968-1978)", en M^a Amparo Tavares Maleval y Laura Tato Fontañá (eds.), *Estudos Galego-Brasileiros 4. Lingua, Literatura, Identidade*, A Coruña, Universidade da Coruña, 2010, pp. 255-276.
- Santos Zas, Margarita, "Los manuscritos de Valle-Inclán, existen", *Anuario Valle-Inclán VIII/ Anales de la Literatura Española Contemporánea*, 32/3 (2008), pp. 5-10.
- Santos Zas, Margarita, "Los manuscritos de Valle-Inclán: el taller del escritor", en Bénédicte Vauthier y Jimena Gamba Corradine (eds.), *Crítica genética y edición de manuscritos hispánicos contemporáneos*, Salamanca, Universidad, 2012, pp. 159-175.
- Santos Zas, Margarita, "Editar a Valle-Inclán: del manuscrito al impreso", en Ermitas Peñas Varela (ed.), *Perspectivas críticas para la edición de textos de literatura española*, Santiago, Servizo de Publicacións da Universidade, 2013, pp. 275-312.
- Serrano Alonso, Javier y de Juan Bolufer, Amparo, *Bibliografía general de Ramón del Valle-Inclán*, Santiago, Servizo de Publicacións da Universidade, 1995.
- Valle-Inclán, Ramón del, *Luces de bohemia. Esperpento* (1924), ed. Alonso Zamora Vicente, Madrid, Espasa-Calpe, Clásicos Castellanos, 1993.
- Vauthier, Bénédicte, *Juan Goytisolo, Paisajes después de la batalla. Preliminares y estudio de crítica genética*, Salamanca, Ediciones de la Universidad, 2012.
- Vauthier, Bénédicte y Gamba Corradine, Jimena (eds.), *Crítica genética y edición de manuscritos contemporáneos. Aportaciones a una poética de transición entre estados*, Salamanca, Ediciones Universidad de Salamanca, 2012.

PÁGINAS DE INTERNET

- <<http://www.oxygenxml.com/>>.
- <<http://www.tei-c.org/index.xml>>.
- <<http://www.tei-c.org/Roma/>>.



RESUMEN

A través de esta contribución pretendemos presentar el *Archivo Digital Valle-Inclán*, asociado a la trayectoria y líneas de trabajo del Grupo de Investigación Valle-Inclán de la USC (GIVIUS), desarrolladas a partir de un fondo documental –corpus manuscrito e impreso– cuya complejidad, riqueza y amplitud ha conducido a la construcción de dicho Archivo, cuyos objetivos, gestión y principales problemáticas se exponen de acuerdo con la tipología de los materiales que conforman dicho fondo documental y bibliográfico, reunido a lo largo de veinte años de investigación del GIVIUS.

Palabras clave: Archivo Digital Valle-Inclán, bases de datos, manuscritos, prensa histórica, herramientas digitales.

ABSTRACT

Our aim is to present *Valle-Inclán's Digital Archive*, associated with the trajectory and work lines of the *Grupo de Investigación Valle-Inclán*, of the University of Santiago de Compostela (GIVIUS), and developed taking as a point of departure a wide, complex and varied documentary collection -here referred to as handwritten and printed *corpus*-. The objectives of this digital tool, its management and main problems found along the process of its creation, are exposed in the following pages, bearing in mind the kind of materials gathered in this documentary and bibliographical archive by the GIVIUS during the last twenty years of research activity.

Keywords: Valle-Inclán's Digital Archive, databases, manuscripts, historical press, digital tools.

De una obra digitalizada a una obra digital: reflexiones en torno al *Diccionario Griego-Español* en línea*

SABINE THULLIER
ILC-CSIC
sabine.thuillier@cchs.csic.es

Un amplio repertorio de obras impresas ha pasado a estar accesible en línea. Podría decirse que el proceso ya resulta banal. Sin embargo, el grado de interés y de éxito de estas empresas no está siempre garantizado. ¿Cómo es posible que una obra ocupe su lugar de pleno derecho en el mundo digital? ¿Cómo la digitalización de obras puede desembocar en una genuina obra digital que suponga un salto cualitativo en la forma de acceso y lectura? Ese salto reside, diremos aquí, en dos de sus rasgos más destacados: el dinamismo y la conectividad.

En este artículo, nos proponemos reflexionar sobre las condiciones que han de darse para que una conversión del papel a la pantalla sea en cierto sentido exitosa. Para ello centraremos nuestra atención en un tipo de texto cuyo aprovechamiento y uso en su formato electrónico es de considerable importancia: los diccionarios. La cuestión que nos interesa debatir aquí es la siguiente: ¿Cómo han de abordarse estos parámetros generales, propios de toda obra digital, en el caso específico de los diccionarios? Para dar una respuesta, provisional, a esta pregunta, vamos a partir de la digitalización del *Diccionario Griego-Español*¹ y reflexionar sobre las posibilidades y limitaciones para su transformación en una obra digital de pleno derecho. Mostraremos que en el caso de los diccionarios convertirse en una obra digital de pleno derecho depende de su conectividad con otros recursos y, sobre todo, de que pueda desarrollarse en diálogo con su corpus de modo dinámico, de tal manera que se haga disponible la información lingüística y lexicográfica para sus potenciales usuarios. Sólo de esta manera podrá convertirse en una herramienta filológica nueva. El proceso no carece de riesgos, pues se produce una cierta paradoja:

* Este estudio ha sido realizado en el marco del proyecto de investigación FFI2010-20236. Agradezco a Juan Rodríguez Somolinos sus aportes y muy especialmente a Jesús Vega su inmensa ayuda.

¹ Más adelante [DGE], <<http://dge.cchs.csic.es/xdge/>>.

a mayor dinamismo y conectividad, quizá una mayor tensión en los valores de rigor y de precisión lexicográficas. El énfasis en los diccionarios como herramientas definidas por sus usuarios potenciales y en la necesidad de aumentar su conectividad y dinamismo no debe hacer perder de vista el horizonte desde el cual proceder a la transformación de un diccionario digitalizado en una obra digital: la propia lexicografía.

Procederemos del siguiente modo. Vamos a presentar los resultados de la digitalización del *DGE* en el marco de la clasificación de los diccionarios digitales propuesta por Sven Tarp (2011 y 2021). El *DGE* en línea es lo que Tarp llama un "caballo más rápido". A continuación, evaluaremos cómo y en qué medida puede convertirse en un tipo de diccionario genuinamente digital. Concluiremos con una discusión de las posibilidades y límites de tal tarea y las consecuencias para los diccionarios de griego antiguo.

1. UNA NUEVA ETAPA DE LA LEXICOGRAFÍA

Todos estaríamos de acuerdo en que las herramientas informáticas desarrolladas en las últimas décadas están contribuyendo a una renovación de la lingüística. Es cierto igualmente que la lexicografía no escapa a este destino. Sin entrar en un debate sobre las relaciones de la lingüística con la lexicografía, no es menos cierto que esta última ocupa un lugar peculiar entre los trabajos lingüísticos. Las investigaciones que emprende quedan reflejadas en un objeto indisolublemente unido a sus potenciales usuarios. No se puede por menos de destacar su carácter artefactual y funcional. Los diccionarios son primariamente *objetos de consulta*.

Este aspecto esencial cobra, si cabe, una importancia decisiva en la configuración de los diccionarios digitales. Por eso mismo, una de las principales preocupaciones ha de ser la de lograr una mayor adaptación a las necesidades reales de los usuarios o, incluso, a potencialidades que están aún por descubrir. Este es un punto central de algunas teorías lexicográficas recientes: la insistencia en las necesidades de los usuarios constituye el rasgo común al estudio de los diccionarios en línea. Se puede decir incluso que su grado de adaptabilidad condiciona su grado de "digitalidad".

Una de las primeras preguntas que cabe hacerse concierne, pues, a las implicaciones teóricas y prácticas de este cambio de soporte para los diccionarios. ¿Qué significa el paso de una lexicografía de papel a una lexicografía electrónica? En primer lugar, es preciso señalar que la distinción no se ha de hacer de forma puramente cuantitativa; involucra esencialmente cambios cualitativos dependientes del modo en que son aprovechadas las funcionalidades de los ordenadores y las tecnologías de la información. Implica más datos, más flexibilidad y adaptabilidad, una mayor manejabilidad, más accesibilidad y una mayor interconexión con fuentes externas.

Si se recuerda, los inicios de la lexicografía electrónica estaban más bien destinados a producir herramientas que fueran legibles para las propias máquinas. La digitalización del griego antiguo es un buen ejemplo de ello. A mediados de los años 90, el proyecto Perseus² emprendió las tareas de digitalización

² <<http://www.perseus.tufts.edu/hopper/>>.

de la edición de 1940 del diccionario de referencia *A Greek English Lexicon*³, con el fin de enriquecer y controlar la base léxica que necesitaba su analizador morfológico. El proyecto llevó a cabo una versión etiquetada de la obra, que es desde entonces de libre consulta en línea y ampliamente usada por otros proyectos⁴. Perseus ha conseguido formalizar un recurso tradicional que, sin embargo, como señalaremos más adelante, está mal adaptado a un entorno electrónico.

Es en este contexto en el que se va a situar nuestra reflexión, el de los avances que puede significar para un diccionario griego proseguir su conversión en obra genuinamente digital. También aquí es necesario insistir en un cambio de énfasis, que no se limite al aprovechamiento "automático" de las informaciones léxicas sino que tenga en cuenta a los usuarios y sus necesidades. Y aquí las peculiaridades del usuario de los diccionarios del griego antiguo aparecen en primer plano. Los usuarios de estas herramientas léxicas no responden a necesidades de *producción* de texto (salvo quizá en el caso de aquellos lingüistas que se sirvan de los diccionarios para producir elementos ficticios para la contrastación de sus hipótesis), sino a necesidades ligadas a la *recepción*, como por ejemplo la traducción de textos. Los diccionarios son objetos de uso cotidiano entre los helenistas, que les atribuyen una cierta autoridad, quizá mayor que en otros casos.

Por eso, es necesario reflexionar sobre qué tipo de tareas de digitalización se ha de emprender con los diccionarios de griego antiguo y, ante todo, qué tipo de obra digital es recomendable. La experiencia en la digitalización del *DGE* puede ser aquí de ayuda. Por ello, es conveniente en primer lugar situarlo entre las categorías de diccionarios que uno puede encontrar en línea. Para ello, vamos a utilizar la clasificación propuesta por Sven Tarp (2011: 58-62 y 2012: 257-259).

Tarp clasifica los diccionarios en línea en cuatro grandes categorías: copiones, caballos más rápidos, modelo Ford-T y Rolls Royce. A pesar de lo que pudiera parecer, la clasificación no responde únicamente a criterios de mayor velocidad en el tratamiento y en el acceso a la información.

- Los "copiones" son copias fotografiadas o escaneadas de diccionarios impresos ya existentes. No existen muchos ejemplos en la lexicografía griega (por lo menos moderna). Citaremos el Bailly *Abrégé*⁵. Su utilidad es escasa, su dinamismo es nulo y, a menudo, resultan menos prácticos que su versión en papel.
- Los "caballos más rápidos" proporcionan una visualización estática en pantalla de los artículos del diccionario. Están modelados a partir de los artículos correspondientes en los diccionarios impresos. Puede decirse que la mayor parte de los diccionarios actualmente en línea son de este tipo y suponen una mejora sustancial por lo que respecta a un acceso más rápido a los datos mediante el uso de enlaces y cadenas de búsqueda. Las varias versiones electrónicas del LSJ corresponden a esta categoría.

³ En adelante [LSJ].

⁴ Entre otros: <<http://philolog.us/>>, <<http://perseus.uchicago.edu/Reference/LSJ.html>> y <<http://stephanus.tlg.uci.edu/lsj/>>.

⁵ <<http://home.scarlet.be/tabularium/bailly/>>.

- Los “modelos Ford-T” ofrecen ya artículos dinámicos con datos dinámicos y adaptados a las distintas funciones que pueden desempeñar los diccionarios. Uno de sus rasgos es el hecho de que ofrecen la capacidad de *reutilizar* datos ya existentes. Logeion, un proyecto en pleno desarrollo que proporciona una búsqueda simultánea en varios diccionarios de latín y griego presentes en la biblioteca digital de Perseus⁶, es quizás el único caso de la lexicografía griega que podría pertenecer en ciertos aspectos a esta categoría.
- Los “Rolls Royce” son concebidos como herramientas lexicográficas avanzadas e individualizadas. En ellas se busca combinar un acceso selectivo a los datos incluidos en una base de datos con la búsqueda en internet (y en corpus existentes). La funcionalidad principal de estos “diccionarios” reside en recrear y re-presentar los datos relevantes en soluciones dinámicas para los usuarios. Tarp comenta que no parece existir aún ninguna herramienta de esta categoría y lo que ofrece la lexicografía digital del griego antiguo no le puede contradecir.

De estos dos últimos tipos de diccionario se puede decir que son obras lexicográficas genuinamente electrónicas, mientras que los dos primeros pertenecen aún a una lexicografía de papel. ¿En qué categoría incluir el *DGE* en línea? Como veremos, su realidad lo sitúa entre los caballos más rápidos; su ambición es avanzar a un modelo Ford-T. Pero ¿por qué no hacer de él un Rolls Royce de los diccionarios en línea? A ello responderemos en la última sección.

2. EL *DGE* EN LÍNEA: UN CABALLO MÁS RÁPIDO

2.1. El proyecto editorial y su digitalización

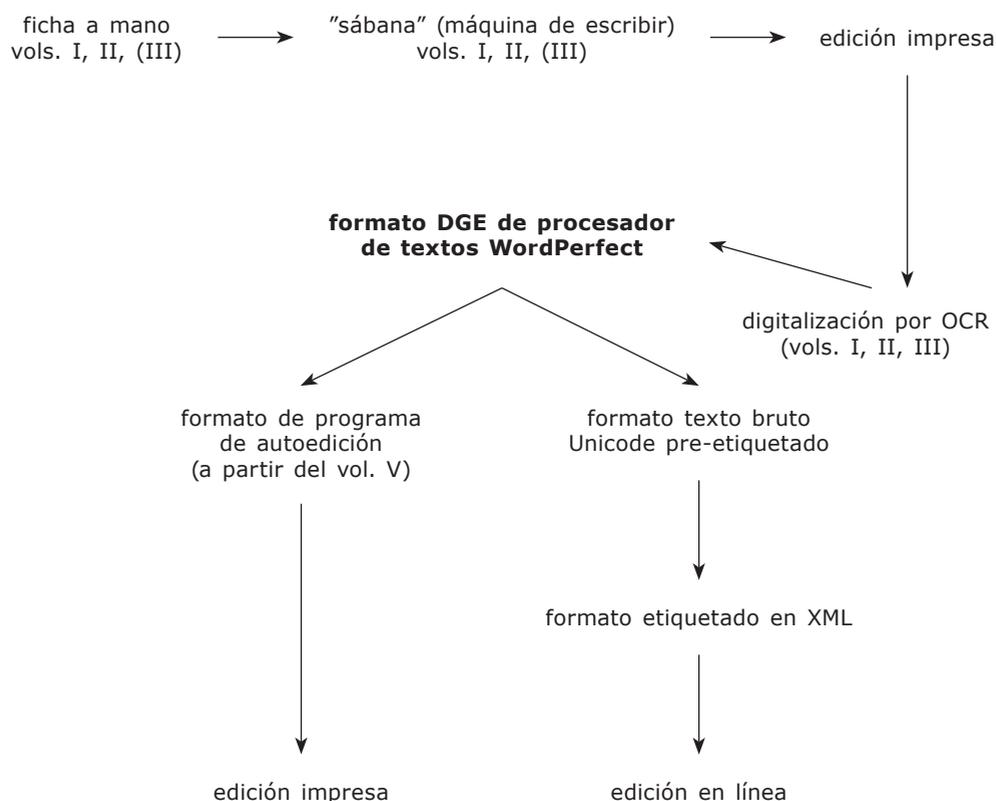
Pocos son los diccionarios del griego antiguo que tengan una versión digital. Recientemente se ha sumado al ya conocido LSJ la parte del *DGE* publicada hasta el momento. A continuación, vamos a presentar este diccionario en el marco de la clasificación de Tarp de los diccionarios electrónicos. Sin duda, por el momento cabe decir que no se trata sino de un “caballo más rápido” que tiene la pretensión de convertirse en un verdadero modelo de Ford-T.

El *Diccionario Griego-Español*, iniciado a finales de los años sesenta y dirigido por el profesor F.R. Adrados, es un diccionario bilingüe unidireccional (griego antiguo – español). Se trata de una obra en curso de elaboración que se publica volumen a volumen desde 1980; los siete volúmenes ya editados cubren la sección alfabética: *ἀ-ἐξαιος* (59.329 entradas). Su ambición es perseguir la larga tradición de la lexicografía griega y contribuir a su renovación, en particular actualizando y enriqueciendo su predecesor inmediato y modelo, el LSJ. El léxico de la lengua griega está analizado de la manera más exhaustiva posible: el corpus despojado comprende todos los textos literarios y documentales transmitidos (incluso el corpus papirológico y epigráfico) y cubre un arco cronológico de más de 1.500 años, desde Homero a los últimos papiros escritos en Egipto en el s. VIII d.C., e incluye la literatura cristiana, los nombres propios, así como el griego micénico. El *DGE*, partiendo lógicamente del LSJ, pretende

⁶ <<http://logeion.uchicago.edu/>>.

reelaborar desde cero la organización y estructura interna de los artículos a partir de la lengua de salida, el español. Pretende establecer en qué circunstancias contextuales (lingüísticas o extralingüísticas) una palabra griega puede o debe traducirse por tal o cual palabra española, precisada o matizada por las explicaciones pertinentes y, eventualmente, los contextos griegos de las citas y sus traducciones⁷.

El proyecto de digitalización del *DGE* se emprende con el objetivo de revalorizar un texto regular y fuertemente estructurado. No es este el lugar de describir cómo se ha llevado a cabo todo este proceso. A modo de resumen, el siguiente esquema presenta los distintos soportes y formatos en que ha viajado el *DGE* desde sus inicios hasta llegar a una versión en HTML:



Por su importancia en este proceso, merece la pena recordar que el proyecto ha procedido a llevar a cabo *LMPG*⁸, edición en línea de una publicación aneja al *DGE*, que ha servido como banco de pruebas. El *Léxico de magia* y

⁷ Para más información sobre la historia y las características del *DGE* remitimos al Prólogo del vol. I y a la presentación en línea: <<http://dge.cchs.csic.es/hist>>.

⁸ <<http://dge.cchs.csic.es/lmpg/>>.

*religión en los papiros mágicos griegos*⁹ sólo incluye 2.627 entradas y analiza semánticamente un corpus homogéneo y restringido, el de los papiros mágicos griegos del s. II a.C. hasta el s. V d.C. Su interés para esta empresa residía precisamente en el hecho de que sigue de modo muy estricto la estructura del *DGE*.

Como el primer objetivo de este proceso era “meramente” obtener una presentación HTML del *DGE*, el tipo de informaciones marcadas ha estado codificado por una consecución del objetivo en tiempo razonable. Para ello se ha diseñado un esquema específico conforme a la *Text Encoding Initiative*¹⁰.

2.2. El *DGE* en línea: estado actual

Obtener una presentación en HTML no significa, en principio, más que un cambio de soporte, un cambio que no obstante supone ventajas sustanciales de legibilidad, aunque el acercamiento sea aún muy similar a la consulta de artículos en papel: buscar lemas, leer artículos. Pero el acceso al texto se hace más fluido y empieza ya a romper la rigidez del papel. Señalaremos algunos de los avances que sugieren un mejor aprovechamiento de potencialidades digitales de la tarea ya realizada:

- La navegación en el diccionario. Sin duda, el *DGE* digitalizado es un caballo más rápido, que permite un mejor y más veloz acceso a la información lexicográfica. Para ello se ha implementado una potente herramienta de transliteración (elemento muy útil para el griego) así como una herramienta de construcción de listados inversos, que podría ser de interés para estudios morfológicos, literarios, etcétera.
- La visualización de los artículos mejora sustancialmente en relación al diccionario en papel, pues se ha insistido en una visualización *estructurada*, basada en mayor espacio entre los apartados repartidos en párrafos indentados. Esto ha significado un cierto valor lexicográfico añadido que pone de relieve la composición muy jerarquizada de los artículos.

La descripción de Tarp (2011: 60) de los caballos rápidos se ajusta perfectamente a los resultados obtenidos: “The result of this restrictive use of the technological possibilities is faster lexicographical horses, where the data included are still organized in traditional and static articles, which are completely modelled on the corresponding articles in printed dictionaries”. Cabe por tanto decir que, de momento, abandonar el papel no se ha traducido en la generación de un objeto digital nuevo. De hecho, este no era el primer objetivo de la digitalización iniciada, más preocupada por mejorar la visibilidad y la accesibilidad a datos lexicográficos básicos. Sin embargo, estos primeros pasos traducen a partir de ahora una toma de conciencia de las nuevas posibilidades ofrecidas por el diccionario. Sin duda, la codificación realizada permite ya implementar funciones más avanzadas de consulta y de búsqueda. Pero sigue siendo un reto avanzar en la conectividad entre los propios recursos del diccionario y abrir también la puerta a nuevas conexiones con recursos ajenos.

⁹ Luis Muñoz Delgado (2001).

¹⁰ <<http://www.tei-c.org/>>. Para más información sobre la codificación del *DGE*, cf. Glorieux-Thuillier (2010).

3. POSIBILIDADES Y LÍMITES PARA UNA MAYOR "DIGITALIDAD"

"If there is one single feature likely to be applicable to all E[lectronic] D[ictionary]s of the future, it is that they will stop functioning as stand-alone products"¹¹.

La sección anterior ha descrito las características que hacen del *DGE* en línea, en su versión actualmente accesible, un diccionario electrónico del tipo que Tarp califica de "caballo más rápido". Es claro que, de momento, su salida del papel no ha logrado producir un nuevo objeto. Pero nos podemos preguntar cuáles son las posibilidades de que un diccionario de este tipo pase a engrosar las filas de la categoría de Ford-T, una categoría que, quiero recordar, se definiría por *artículos dinámicos* que contienen a su vez *datos dinámicos*. Más allá de un recurso lexicográfico digitalizado, meramente convertido de un formato a otro, el reto fundamental es perseguir el tratamiento de los datos y lograr así un recurso digital dinámico, enlazable y compartible. Vamos por consiguiente a proponer pistas para desarrollar la conectividad y el dinamismo de varios tipos de datos contenidos en el *DGE*, tanto los datos internos al propio proyecto como otros procedentes de otras fuentes. En un segundo momento, explicaremos los límites previsibles a desarrollos en esta línea y las razones por las que una obra como el *DGE* no puede pretender alcanzar la categoría de Rolls Royce. Al final propondremos algunas reflexiones sobre las consecuencias para la evolución futura del *DGE*.

3.1. Conectarse a sí mismo: dinamizar los datos internos

Dinamizar los campos. Ya hemos dicho que las funcionalidades actuales del *DGE* en línea no proporcionan realmente accesos novedosos a los datos. Es, por tanto, preciso conseguir que se puedan llevar a cabo múltiples búsquedas en la base de datos. Este nuevo tipo de navegación, que rompería aún más la linealidad del texto, podría así alcanzar un cierto estadio avanzado de sofisticación según sea el grado de granularidad de la codificación de la fuente y la implementación de un motor de búsquedas adecuado. Esto podría producir una especie de reorganización de los datos y construir así nuevas redes a través de distintos artículos. Los resultados obtenidos en los modos de consulta avanzada del *LMPG* son ya convincentes (búsqueda a partir del corpus citado¹², traducciones españolas, marcas de uso, etc.¹³) y es preciso adaptar este sistema de búsqueda al *DGE*. Será posible entonces visualizar las informaciones de un modo que es imposible en el LSJ, cuya fuente XML está marcada de modo mucho menos fino.

Del mismo modo, se obtendrá un valor añadido de primera magnitud al extender este dinamismo a otros datos esenciales del diccionario y que están almacenados en otros lugares. He aquí algunos ejemplos:

Conectar las citas con su información bibliográfica. Los siete volúmenes en línea del *DGE* contienen 370.000 citas. Recordemos que se trata de una descripción lexicográfica fundada sobre el corpus y que toda acepción, todo matiz semántico, debe estar documentado por una o, muy a menudo, varias

¹¹ De Schryver (2003: 189).

¹² <<http://dge.cchs.csic.es/lmpg/?textos=>>.

¹³ <<http://dge.cchs.csic.es/lmpg/busqueda>>.

referencias al corpus. Más que desempeñar un papel de ilustración, el reparto de las referencias textuales constituye de hecho la base de organización de los artículos. Como es común en los diccionarios de lenguas antiguas impresos, el sistema de abreviaturas empleado para estas referencias es particularmente complejo y denso, y resulta tal vez poco intuitivo para muchos lectores. Dada la importancia de este campo, parece necesario enlazar de manera más fluida estas abreviaturas a su forma completa recogida en las listas bibliográficas informatizadas pero aisladas en otra parte del servidor¹⁴. Estas listas detallan todas las ediciones de referencia que utiliza el *DGE* y están en permanente proceso de actualización; representan una fuente de primer orden para quien quiere disponer de una información bibliográfica bastante completa sobre el corpus griego editado. Dinamizar la citas del diccionario y conectarlas con estas listas, además de hacerlas más legibles, abrirá la posibilidad, entre otras, de navegar a través el *DGE* a partir del corpus analizado, crear concordancias, etcétera.

Documentos léxicos. Existen muchos otros datos, muy valiosos, desarrollados en el proyecto filológico alrededor de la elaboración del *DGE*. Pensamos en primer lugar en el *Repertorio Bibliográfico de la Lexicografía Griega*¹⁵. Bajo el artículo aparecerían las referencias bibliográficas a estudios léxicos sobre la palabra que hay en RBLG (algunas de ellas además quizá también accesibles en Internet). Sin duda, la literatura secundaria ocupa un lugar muy importante en el trabajo de los lexicógrafos del *DGE*. Esta conectividad sería un modo de hacer visible el trabajo filológico que está en el origen de los artículos y proporcionaría también una ayuda innegable a ulteriores investigaciones¹⁶.

Se podrá también incorporar al *DGE* en línea los muy numerosos *addenda et corrigenda* que se acumulan en los archivos: palabras nuevas, palabras fantasma, nuevas acepciones, corrección de errores, etc. Los *addenda et corrigenda* se presentarían integrados en el *DGE* pero de un modo que los haga claramente identificables, esto es, manteniendo el original y el suplemento.

3.2. Conectarse con datos externos

Los proyectos digitales de la comunidad de estudios clásicos están en plena expansión. La cantidad de información relacionada con datos sobre el mundo clásico disponibles en línea va aumentando día a día, y es identificable un deseo de formular consultas capaces de cruzar estos diferentes almacenes de datos. Con toda seguridad, el *DGE* tiene un papel que desempeñar en esta comunidad digital.

Los diferentes tipos de recursos que se van a proponer aquí se reparten según su orden de proximidad con el *DGE*. Es obvio que esta lista podría quizá extenderse más lejos, a todo lo que contiene internet: "From a sound metalexigraphic point of view this is hard to accept, yet the Internet can definitely be used as the largest available collection of electronic data [...], with many hundreds of dedicated corpora also accessible" (de Schryver 2003: 158)

Otros recursos lexicográficos. La conexión con proyectos similares permite sin duda un aprovechamiento conjunto de los recursos y ofrece a los usuarios

¹⁴ <<http://dge.cchs.csic.es/lst/lst-int.htm>>.

¹⁵ RBLG: <<http://dge.cchs.csic.es/blg/blg-s.htm>>.

¹⁶ Cf. Lee (2010: 126 y 134).

la posibilidad de contrastar la información léxica proporcionada por varios diccionarios. Aparte del caso obvio de conectar los artículos del *DGE* con los del LSJ, pensamos igualmente en otros diccionarios de griego antiguo nativamente digitales, que saldrán a la luz próximamente, como el Cambridge Greek Lexicon¹⁷ y el dicciogriego¹⁸. Otro posible tipo de intercambio tiene más que ver con la lexicografía computacional. La fuente léxica etiquetada del *DGE* puede contribuir al desarrollo de bases de datos léxicos existentes. Esto permitiría actualizar y corregir los lematizadores disponibles (Perseus, TLG). El *DGE* contiene en efecto un número considerable de lemas que faltan en el LSJ. A cambio, el *DGE* debe a su vez implementar estas herramientas lingüísticas, para permitir por ejemplo búsquedas lematizadas de las palabras griegas, función elemental de cualquier diccionario electrónico.

"Re-conectar" un diccionario a su corpus¹⁹. Conviene aquí insistir sobre la situación excepcional del corpus literario y documental de la lengua griega antigua: una parte considerable de los textos griegos está disponible en formato digital y en gran medida lematizado. Entre el número creciente de proyectos editoriales, nos gustaría destacar el importantísimo *Thesaurus Linguae Graecae*²⁰ que ha recopilado y digitalizado la mayoría de textos literarios griegos, desde Homero hasta 1453²¹ y la biblioteca digital de Perseus, de mucho menor tamaño²² pero que está considerada en el ámbito de las humanidades digitales como un recurso de referencia por la interactividad de los datos almacenados. Por lo que se refiere al corpus no literario, es preciso citar para los papiros el Duke Databank of Documentary Papyri²³, y para las inscripciones tanto el Packard Humanities Institute Searchable Greek Inscriptions²⁴ como el *Supplementum Epigraphicum Graecum*²⁵.

Dada esta abundancia de fuentes textuales en formato digital, la lengua griega está en una situación particularmente favorable: casi cualquier hipótesis lingüística se puede comprobar en la casi totalidad de las ocurrencias conservadas del léxico. Así, se puede entender sin dificultad que la interconexión del corpus con su descripción lexicográfica resulte ser una importante clave de la lexicografía digital del griego antiguo. Por eso, la afirmación de Toma Tasovac (2010), "we cannot think of dictionaries any more without thinking about digital libraries and the status which electronic texts have in them", nos parece aplicarse de manera muy oportuna al ámbito helenista. Esta interconexión ya se ha hecho efectiva en Perseus y en el TLG. Ambos integran en su colección diccionarios, entre los que se cuenta el LSJ. Esta interoperabilidad corpus-diccionario constituye el mayor desafío del actual desarrollo digital del *DGE* en línea. Uno de los mayores provechos residiría en que los usuarios pudieran acceder de manera rápida y fácil a los numerosos textos a los cuales *DGE* sólo hace referencia, sin proporcionar contexto.

¹⁷ <http://www.classics.cam.ac.uk/faculty/research_groups_and_societies/greek_lexicon/>.

¹⁸ <<https://sites.google.com/site/dicciogriego/>>.

¹⁹ Cf. de Schryver (2003: 167-168) y Rundell (2012: 78-77).

²⁰ <<http://stephanus.tlg.uci.edu/>>.

²¹ "105 million words from over 10,000 works associated with 4,000 authors".

²² Acerca de 10 millones de palabras griegas.

²³ <<http://papyri.info/>>.

²⁴ <<http://epigraphy.packhum.org/inscriptions/>>.

²⁵ <<http://hum.leiden.edu/history/research/projects-umw/seg.html>>.

No obstante llevar a cabo un sistema hipertextual adecuado entre el diccionario y los corpus es una empresa delicada que requiere tanto un trabajo colaborativo amplio como gran precaución. Por un lado, existe la dificultad de normalizar los identificadores de las citas, dada la alta variación de formas de citar los textos; alcanzar un marcado operativo al nivel de la referencia exacta de una cita es una tarea costosa. Por otro lado, existe el riesgo de remitir a una edición diferente (a menudo más antigua) de la adoptada por el diccionario. Eso conlleva la posibilidad de tener lecturas diferentes del mismo pasaje y por lo tanto menoscabar el atento análisis del lexicógrafo. Seguramente, se puede encontrar toda una serie de soluciones para cruzar, tecnológica y científicamente, la barrera de estas variaciones textuales y editoriales²⁶. Cabe señalar, por otra parte, que no todos los recursos mencionados están en acceso abierto, y esto puede naturalmente dificultar los protocolos de intercambio de datos con algunos proyectos.

Datos extra-lingüísticos. El *DGE* no tiene intenciones enciclopédicas propiamente dichas. Sin embargo, a diferencia del LSJ, incluye nombres propios (más de 10.000). Sin seguir los criterios de exhaustividad aplicados al resto del léxico, sus motivaciones quedan todavía a un nivel lingüístico y filológico²⁷. En este dominio, la comunidad de estudios clásicos es especialmente activa y se puede contar con una cantidad de recursos que recopilan datos de carácter enciclopédico sobre el mundo antiguo. Son recursos geográficos (el proyecto Pleiades²⁸ por ejemplo), biográficos y onomásticos (como el gran *Lexicon of Greek Personal Names*²⁹), iconográficos (como el *Lexicon Iconographicum Mythologiae Classicae*³⁰), sin olvidar todos los metadatos disponibles para los corpus documentales, sobre todo papirológicos³¹. Conectar el *DGE* con recursos de este tipo permitiría facilitar una justa ampliación de su modesta parte extra-lingüística³².

3.3. Límites y brechas

En lo dicho hasta ahora, hemos intentado proponer diferentes pistas para dinamizar el *DGE* en línea dentro de un marco que le sería, por así decir, favorable y contribuiría a proporcionar accesos a los datos y herramientas digitales novedosos y útiles para la investigación filológica. El grado de factibilidad técnica de esta tarea es variable y depende de los tipos de interconexiones previstas. Pero nos parece que el límite más serio a una dinamización más amplia del

²⁶ Conviene mencionar los esfuerzos de estandarización desarrollados por el Canonical Text Service, recientemente seguidos por el Perseus Catalog, y también por la comunidad papirológica.

²⁷ "Los nombres propios son tan signos lingüísticos como los comunes, que a veces conservan o presentan lexemas no documentadas en otras clases de palabras, que un diccionario es un instrumento de trabajo que debe resolver dudas a un traductor (y no siempre es fácil saber ante un nombre propio si se trata de un mes, de persona, o de lugar, lo cual influiría, por ejemplo, en la traducción que rige al nombre propio), etc.", Prólogo del vol. I del *DGE*, p. XXXV.

²⁸ <<http://pleiades.stoa.org/>>.

²⁹ <<http://www.lgpn.ox.ac.uk/online/index.html>>.

³⁰ <<http://www.limc-france.fr/>>.

³¹ Entre otros, <<http://www.papyri.info/>>.

³² Cf. Rundell (2012: 78): "the boundaries between dictionary and encyclopedia (always problematic) is increasingly irrelevant".

contenido de este diccionario procede sobre todo de cuestiones teóricas y científicas. Hemos hablado más arriba de las precauciones que hay que tomar a la hora de vincular las referencias de los textos citados por el *DGE* con los corpus electrónicos que proponen estos "mismos" textos, con el consiguiente riesgo de alterar la precisión de los datos. Por lo tanto, varios puntos contradictorios y delicadas preguntas surgen inevitablemente acerca del grado de adecuación de este tipo de recursos –a pesar de todo diseñados de una forma tradicional– a un ámbito digital "hiperconectado" y altamente modulable y adaptable según las necesidades de los usuarios, con el fin de lograr alcanzar la categoría de Rolls Royce descrita por Tarp.

Evocaremos brevemente varios aspectos que sugieren estas limitaciones; profundizar en el asunto que está en juego desbordaría el propósito de estas páginas. En primer lugar, si bien las capacidades de almacenamiento de datos internos y enlaces a fuentes externas no parecen tener ya límites físicos, no se puede abrumar al usuario con información que podría resultar disparatada. Lo repetimos, el diccionario es (y tal vez aún más en su forma electrónica) una herramienta de consulta puntual, que secunda en la mayoría de los casos otra actividad principal. Existe la necesidad de una mayor claridad, pertinencia, oportunidad, con la menor distracción posible³³. ¿Hasta qué punto estos requisitos son compatibles con los sueños de una interconexión *ad libitum*? La primera medida consistiría en elaborar un motor de búsqueda capaz de adaptarse a las necesidades del usuario y proponer esta "pos-navegación" como una elección discreta y jerarquizar así los vínculos según su pertinencia, siempre incluyendo una prevención sobre el valor filológico de lo que puede ofrecerse. No se trata sólo de la cuestión trivial de cómo se debe presentar la información sino también de responder a las exigencias de exactitud que esperan de las explicaciones lexicográficas los usuarios, que son en el caso que discutimos aquí, filólogos y más generalmente clasicistas.

Finalmente, ante esa paradoja amenazante, a saber, que al aumentar su conectividad el diccionario corre el riesgo de perder en parte su rigor científico, un proyecto lexicográfico como el *DGE*, procedente del papel, tiene que reasentar sus ambiciones para no sólo soslayar esta paradoja sino también para aprovechar más a fondo el cambio radical que implica la era digital para la lexicografía. El *DGE* se encuentra ahora de cierta manera a caballo entre el papel y la pantalla, tanto por lo que respecta a su parte ya publicada y digitalizada como por lo que respecta a lo que queda por elaborar. Hemos intentado a lo largo de estas páginas sugerir unas pistas para que el contenido actual de *DGE* en línea, a saber siete volúmenes impresos digitalizados, pueda en un futuro ganar un lugar de pleno derecho en el seno de los diccionarios digitales. Estos retos podrán quizás servir para encarar con mayores garantías la necesaria adaptación de sus pretensiones y de sus métodos, para proseguir la redacción de la obra al tiempo que se desembaraza definitivamente del papel.

BIBLIOGRAFÍA

A Greek-English Lexicon, H.G. Liddell, R. Scott, R. y H.S. Jones (eds.), Oxford, Clarendon, 1940 (9.ª ed.).

³³ Cf. de Schryver (2003: 189).

- De Schryver, Gilles-Maurice, "Lexicographers' Dreams in the Electronic-Dictionary Age", *International Journal of Lexicography*, 16 (2) (2003), pp. 143-199.
- Diccionario Griego-Español*, Francisco Rodríguez Adrados et al. (eds.), Madrid, Consejo Superior de Investigaciones Científicas, 1980-.
- Glorieux, Frédéric y Thuillier, Sabine, "Grec ancien, latin médiéval, balisage comparé de deux dictionnaires, vers des ressources linguistiques", *Archivum Latinitatis Medii Aevi*, 68 (2010), pp. 161-181.
- Lee, John A.L., "Releasing Liddell-Scott-Jones from its past", en Christopher Stray (ed.), *Classical Dictionaries: Past, Present and Future*, London, Duckworth, 2010, pp. 119-137.
- Rundell, Michael, "'It works in practice but will it work in theory?' The uneasy relationship between lexicography and matters theoretical", en R.V. Fjeld y J.M. Torjusen (eds.), *Proceedings of the 15th EURALEX Congress*, Oslo, University of Oslo, 2012, pp. 47-92.
- Tarp, Sven, "Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction", en Pedro A. Fuertes-Olivera y Henning Bergenholtz (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography*, Londres/Nueva York, Continuum, 2011, pp. 54-70.
- Tarp, Sven, "Online dictionaries: today and tomorrow", *Lexicographica*, 28 (2012), pp. 253-268.
- Tasovac, Toma, "Reimagining the Dictionary, or Why Lexicography Needs Digital Humanities", en *Digital Humanities*, 2010, <<http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-883.html>> [29/09/2013].



RESUMEN

Un número considerable de trabajos filológicos impresos han pasado a ser consultables en línea. No obstante, puede afirmarse que este proceso de digitalización no siempre se traduce en la generación de genuinas obras digitales. Como lo indica Sven Tarp, esto es especialmente significativo en el caso de los diccionarios electrónicos, que han de dejar de ser objetos estáticos de consulta para reconceptualizarse como textos dinámicos y diseñados para conectarse con proyectos colaborativos más amplios. ¿Cuáles pueden ser los desafíos a los que se enfrentan los diccionarios para convertirse en recursos propios del ámbito digital? Para abordar esta cuestión, este estudio pretende describir las estrategias que están siendo implementadas en el *Diccionario Griego-Español* recientemente digitalizado (<http://dge.cchs.csic.es/xdge/>). Sugerimos unos caminos que se pueden seguir para dinamizar su contenido y interactuar con fuentes externas (en particular con los corpus en línea). El cambio de formato ha sumido *de facto* el DGE en un entorno cuyos desarrollos en la filología clásica son ya considerables; su papel activo en las aportaciones de la lexicografía en las humanidades digitales dependerá de que se convierta en una obra digital de pleno derecho.

Palabras clave: Lexicografía digital, griego antiguo, corpus.

ABSTRACT

A high number of printed philological works have become available online. However, it can be argued that this process of digitization is not always the guarantee of the creation of genuine digital works. As indicated by Sven Tarp, this issue is especially relevant in the case of electronic dictionaries that should not be static objects of consultation anymore, but should be reconsidered as dynamic texts featured to connect with broader collaborative projects. What might be the challenges that dictionaries face to convert themselves in resources actually belonging to the digital world? To address this question, this study describes the strategies that are being implemented in the newly digitized *Diccionario*

Griego-Español (<http://dge.cchs.csic.es/xdge/>). We suggest some ways to follow in order to make its content more dynamic and to interact with external sources (in particular with online corpora). The change of format has plunged *de facto* the *DGE* in an environment whose developments in classical philology are already considerable, its active role to determine how lexicography can contribute to digital humanities remains in its ability to become a digital work in its own right.

Keywords: Digital lexicography, ancient Greek, corpus.

Post Scriptum: Archivo Digital de Escritura Cotidiana

GAEL VAAMONDE
ANA LUÍSA COSTA
RITA MARQUILHAS
CLARA PINTO
FERNANDA PRATAS

CLUL, Universidade de Lisboa
gaelvmnd@gmail.com

1. INTRODUCCIÓN

El proyecto de investigación *P.S. Post Scriptum: Archivo Digital de Escritura Cotidiana en Portugal y España en la Edad Moderna*¹ (en adelante, *Post Scriptum*) tiene como finalidad recoger y publicar cartas particulares escritas en portugués y en español durante la Edad Moderna (desde el s. XVI hasta el primer tercio del s. XIX) por personas pertenecientes a diferentes estratos sociales. Estas cartas, en su mayoría inéditas, sobrevivieron excepcionalmente al cruzarse las vidas de sus autores con los medios de persecución tanto de la Inquisición como de distintos tribunales civiles y eclesiásticos, instituciones que solían hacer uso de la correspondencia privada como prueba de los delitos que estaban siendo juzgados. Los documentos, que forman parte de procesos judiciales, vienen generalmente acompañados de 'interrogatorios sociológicos' llevados a cabo por inquisidores y jueces varios, lo que permite a los investigadores una contextualización adecuada de las relaciones interpersonales en las sociedades tradicionales. Estas fuentes escritas suelen presentar una retórica casi oral, tematizando asuntos cotidianos que hasta ahora no se habían estudiado más que a partir de casos aislados.

A partir de estas premisas, el objetivo fundamental de *Post Scriptum* es la creación de una base de datos electrónica formada por 7.000 de estas cartas (3.500 para cada lengua) que sirva como herramienta de trabajo para diferentes estudios humanísticos, principalmente para aquellos relacionados más

¹ El proyecto *Post Scriptum* está siendo financiado por el Consejo Europeo de Investigación (7FP/ERC Advanced Grant – GA 295562) y se desarrolla actualmente en el Centro de Lingüística da Universidade de Lisboa (CLUL). Además de los firmantes de este texto, en este momento forman parte del equipo investigador Guadalupe Adámez, Sandra Antunes, Catarina Carvalheiro, Tiago Castro, Elisa García, Raíssa Gillier, Mariana Gomes, Ana Leitão, Laura Martínez, Víctor Pampliega, Liliana Romão, Carmen Serrano y Leonor Tavares.

directamente con disciplinas como la historia moderna, la historia cultural, la crítica textual, la lingüística diacrónica y la lingüística de corpus. Por lo tanto, la edición digital en línea de *Post Scriptum* ofrece un conjunto heterogéneo de cartas escritas en diferentes contextos sociales y que obedecen a situaciones comunicativas variadas. En el ámbito de la lingüística diacrónica, la naturaleza dialógica de estos documentos privados permite compensar, en su justa medida, la carencia de fuentes orales. Por un lado, la espontaneidad en las interacciones que se generan a través de la correspondencia privada puede servir como una ventana al discurso cotidiano (Nevalainen y Tanskanen, 2007); por otro lado, un repertorio de cartas propias de contextos informales, producidas por manos poco instruidas y escritas casi como si fuesen habladas, constituye un recurso extraordinario para el estudio fonológico, morfológico y sintáctico de un determinado período histórico. Finalmente, los datos biográficos de individuos anónimos, así como sus formas de vida y sus interrelaciones sociales, suponen un valor de interés indiscutible, tanto desde la perspectiva histórica y cultural como desde la historiografía moderna.

En el presente trabajo, se explicará la metodología utilizada en *Post Scriptum* para la edición digital de los documentos y su disponibilidad en línea; además, se discutirán cuestiones relacionadas con la modernización de los textos y con su posterior etiquetación morfológica y sintáctica. Este artículo está estructurado de la siguiente manera: el siguiente apartado se centra en el proceso de localización y selección de las cartas; en el tercer apartado, se explica el procedimiento para la edición digital de los documentos, mediante la transcripción en formato XML; el cuarto apartado está dedicado al proceso de modernización de los textos y a su posterior etiquetación lingüística; finalmente, en el apartado quinto se alude a la relación entre diferentes tareas de edición; el trabajo se cierra con unas conclusiones generales sobre el proyecto.

2. BÚSQUEDA Y SELECCIÓN DE CARTAS

En un proyecto como el de *Post Scriptum*, donde lo que se pretende es construir y publicar una base de datos, la primera tarea que cobra importancia es la localización y selección de los datos. En la planificación de tareas que se ha trazado a lo largo de este proyecto, se han reservado los dos primeros años casi exclusivamente a la búsqueda de los textos junto a su transcripción en formato digital (ver el apartado siguiente), lo que deja ya entrever la dificultad de esta labor².

Por lo que se refiere a la localización de datos, la recuperación de correspondencia privada de los siglos XVI al XIX se lleva a cabo mediante la consulta *in situ* de archivos históricos, preferentemente, y como ya se apuntó más arriba, de aquellos que cuenten con un repertorio más o menos amplio de fondos judiciales y/o inquisitoriales. Así, para la búsqueda de cartas españolas se han consultado o se están consultando fondos documentales en el Archivo Histórico General (Madrid), el Archivo General de Simancas (Valladolid) y el Archivo del Reino de Galicia (La Coruña), además de otros archivos en diversos lugares de

² Por lo que se refiere a la búsqueda en archivos españoles, la media aproximada es de una carta encontrada por cada 10 procesos consultados; en archivos portugueses, la media es de 1 carta por cada 11 procesos.

la geografía española (Asturias, Cuenca, Guadalajara, Murcia, Orense, Pontevedra y Toledo)³. En cuanto al caso portugués, la mayor parte de documentación judicial e inquisitorial se encuentra centralizada en el Arquivo Nacional da Torre do Tombo (Lisboa); no obstante, también se están teniendo en consideración otros archivos no nacionales, como son los de Évora, Braga y Oporto, así como el archivo de Goa (India).

La consulta de diferentes archivos a lo largo y ancho de toda la Península Ibérica no está pensada sólo como una forma de aumentar las probabilidades de éxito en el hallazgo de cartas, sino que obedece también a la necesidad de controlar una cierta representatividad geográfica y lingüística de los datos. Junto a la tarea de localización, por tanto, se ha adoptado una estrategia de selección de las cartas encontradas. Además, este proceso selectivo se aplica tanto a la variabilidad espacial como a la variabilidad temporal de los textos: para asegurar el equilibrio cronológico del corpus, la previsión de 3.500 cartas por cada lengua se distribuye del siguiente modo: 500 para el siglo XVI –toda vez que la documentación judicial quinientista es difícil de encontrar en archivos–, 1.250 para el siglo XVII, 1.250 para el siglo XVIII, y 500 para el siglo XIX. Las cartas del Post Scriptum acaban en 1834, fecha representativa de los procesos de reforma de la administración que se siguieron al final del Antiguo Régimen.

Finalmente, también se aplica un control de selección en función del tipo de delito al que se vincula la carta. Experiencias previas en la consulta de fondos documentales nos han permitido comprobar que los procesos judiciales asociados a determinados delitos son más propensos a incluir misivas como prueba instrumental (por ejemplo, las acusaciones de bigamia o las de solicitud⁴). Para evitar un desequilibrio en este sentido, en aquellos archivos que albergan una cantidad considerable de documentación se ha procedido de manera selectiva, consultando siempre un porcentaje mínimo de procesos por cada delito existente.

En definitiva, los trabajos de búsqueda y selección de cartas están encaminados a la construcción de un corpus equilibrado y que sea representativo de los intereses de investigación en Post Scriptum: la escritura cotidiana epistolar en español y portugués durante la Edad Moderna.

3. TAREA PALEOGRÁFICA Y EDICIÓN DIGITAL

Una vez que se localiza la carta, el siguiente paso es transcribir el manuscrito, convirtiéndolo a un formato que sea legible por un ordenador. Y esto conlleva una serie de decisiones técnicas y metodológicas.

La codificación de los datos, tanto textuales como extratextuales, se realiza mediante el uso de lenguaje XML. Los ficheros XML son legibles, sin pérdida de información, por todos los procesadores de texto, lo que facilita su conversión para otros formatos y evita problemas de procesamiento electrónico.

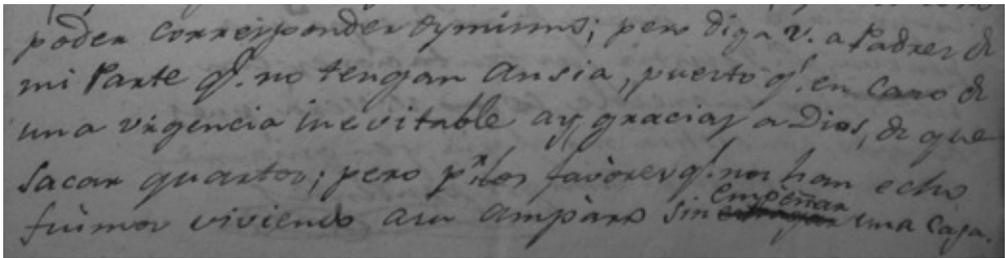
Buscando practicidad y garantías de integración con otros corpus electrónicos de naturaleza similar, se han adoptado los estándares de anotación propuestos

³ En un futuro próximo, se contempla la consulta de fondos en archivos de Barcelona, Granada, Sevilla y Zaragoza.

⁴ El delito de solicitud incluye cualquier tipo de provocación, insinuación o seducción del confesor hacia el penitente.

por el consorcio TEI (*Text Encoding Initiative*), por ser una convención internacional ya consolidada en la edición virtual de fuentes primarias. De todos los proyectos que siguen las directrices TEI, destaca el DALF (*Digital Archive of Letters in Flanders*), dedicado específicamente a la edición epistolográfica. Por consiguiente, la Definición del Tipo de Documento utilizada en Post Scriptum es un instrumento proporcionado por el proyecto DALF (DALF.dtd).

Para la transcripción del manuscrito se ha adoptado una actitud bastante conservadora, lo que da lugar a una edición semipaleográfica del texto original. Tan sólo se ha normalizado la segmentación de palabras y el uso de las grafías "i", "j", "u" y "v". Los cambios de línea, la ortografía, las abreviaturas, los tachones, las correcciones del autor, los accidentes del soporte o la orientación de la escritura, entre otros aspectos, se han respetado en la edición digital, mediante el uso de las etiquetas XML definidas por los proyectos TEI-P4 y DALF. Esto permite ofrecer una edición electrónica del texto manuscrito sin perder rigor filológico. Como muestra de lo dicho, sirva el ejemplo recogido a continuación, en el que las etiquetas "</lb>", "", "<add>" y "<abbr>" permiten marcar cambios de línea, tachones, añadidos autoriales fuera de línea y abreviaturas, respectivamente:



```
poder corresponden oy mismo; pero diga <abbr>u<expan>sted</expan></abbr> a Padres de<lb/> mi Parte
<abbr>q<expan>ue</expan></abbr> no tengan ansia, puesto <abbr>q<expan>ue</expan></abbr> en caso
de<lb/> una vigencia inevitable ay, gracias a Dios, de que<lb/> sacar quartos; pero
<abbr>p<expan>o</expan>r</abbr> los favores <abbr>q<expan>ue</expan></abbr> nos han echo<lb/>
fuimos viviendo a su ampáro sin <del hand="JBC4">entregar</del> <add hand="JBC4"
place="supralinear">empeñar</add> una casa.<pb n="30v"/>
```

Fig. 1. Fragmento de una carta transcrita con etiquetas TEI-XML.

La información que ofrece Post Scriptum de cada manuscrito no se limita a la transcripción semipaleográfica del texto en sí, sino que incluye también diferentes datos de carácter extratextual, desde las características del soporte físico (disposición gráfica del texto, medidas del papel, estado de conservación) hasta la descripción histórica y contextual de la carta. Además, siempre que es posible se facilitan los datos biográficos de los autores y destinatarios de las cartas editadas (nombre, origen, ocupación, religión, estado civil...), los cuales se recopilan en una base de datos independiente.

4. INVESTIGACIÓN LINGÜÍSTICA

El proyecto Post Scriptum se propone enriquecer la investigación en lingüística diacrónica, beneficiándose para ello de los avances que proporcionan actual-

mente disciplinas como la lingüística de corpus o la lingüística computacional. Así, entre los objetivos de Post Scriptum, ocupa un lugar importante la anotación de tipo lingüístico. En principio, se ha contemplado la anotación morfosintáctica de todo el corpus (p. ej., etiquetado de clases de palabras) y la anotación sintáctica de, al menos, una parte importante de los datos.

Hoy en día, existen a disposición del investigador diferentes programas de etiquetación automática que agilizan este tipo de trabajo; no obstante, para obtener de ellos el máximo rendimiento y fiabilidad es deseable que los datos de entrada presenten una ortografía normalizada. Por eso, en Post Scriptum la anotación lingüística está supeditada a la normalización ortográfica previa de los textos que van a ser etiquetados. En el presente apartado se explicará la metodología adoptada para llevar a cabo ambas tareas, normalización y anotación, tanto para el portugués como para el español.

4.1. Normalización ortográfica

Los manuscritos originales de las cartas presentan una gran variabilidad ortográfica. Así, una misma palabra (p. ej., *vergüenza*) puede aparecer escrita de muy diversas formas (p. ej., *berguensa*, *verguensa*, *berguenza*, *vergüenza*, *berguença*, *verguença*...). Esta diversidad, que tiene un interés filológico y lingüístico en sí misma, es respetada escrupulosamente en la edición digital semipaleográfica de Post Scriptum. Sin embargo, a efectos de anotación automática resulta contraproducente, lo que hace necesario un proceso de modernización ortográfica de los textos.

Frente a la edición semipaleográfica, en la edición modernizada se ha normalizado la grafía y la acentuación de todas las formas, se han revisado los signos de puntuación y se ha dividido el texto en párrafos. Además, se han expandido todas las abreviaturas⁵, a excepción de: etc, PD, AD, XPTO y similares. Conviene precisar que las modificaciones realizadas sobre el texto primario se cifien únicamente al nivel ortográfico, por lo que no se eliminó ni se añadió ninguna palabra respecto del contenido original de la carta⁶. Tampoco se ha intervenido sobre el nivel léxico: se han conservado los regionalismos y los arcaísmos léxicos, aunque se han señalado con una marca [*sic*] para posibilitar su recuperación.

El proceso de normalización de los textos se realiza mediante la herramienta eDictor (Faria, Kepler y Sousa, 2010), desarrollada por el grupo de investigación del proyecto Tycho Brahe. Se trata de un programa de interfaz amigable que permite seleccionar la palabra original y editarla de acuerdo con la grafía estándar actual. Además, una ventaja importante que presenta esta herramienta es su compatibilidad con las extensiones TXT, XML y HTML, lo que permite crear archivos de salida en cualquiera de estos formatos. Una vez normalizado el texto, eDictor también es capaz de organizar el contenido en tablas de datos, lo que facilita el visionado de las modificaciones que fueron realizadas en cada nivel de edición (p. ej., formas normalizadas, abreviaturas expandidas y variedades léxicas señaladas).

⁵ En la edición semipaleográfica, la extensión de las abreviaturas aparece marcada con la etiqueta <expan> (ver Fig. 1).

⁶ El texto correspondiente al sobrescrito de la carta, si es que lo hay, queda fuera de la edición normalizada.

El inconveniente de este proceso de edición ortográfica es que ha de realizarse de manera manual, es decir, seleccionando y modificando palabra por palabra todas aquellas formas que sean objeto de normalización. Por eso, en Post Scriptum se está trabajando en un procesamiento semiautomático que agilice esta tarea. De momento, se están haciendo pruebas con la herramienta VARD 2 para el portugués (Hendrickx y Marquilhas, 2011), aunque su aplicación al proyecto todavía se encuentra en fase experimental.

4.2. Anotación lingüística

Los textos electrónicos que constituyen un corpus se pueden presentar de dos formas posibles: no anotados (en su estado original) y anotados (enriquecidos con varios tipos de información lingüística). Como ya se ha mencionado, uno de los objetivos de Post Scriptum es la creación de un corpus anotado de escritura cotidiana, que facilite al usuario la recuperación y el análisis de toda aquella información lingüística contenida en los textos. Actualmente, en Post Scriptum esta tarea incluye al menos la consideración de dos niveles de información: el etiquetado por clases de palabras y la anotación sintáctica. En ambos casos, la anotación se realiza sobre la edición modernizada de los textos y mediante el uso de diferentes programas informáticos, según la lengua de entrada. El proceso de anotación es siempre semiautomático, esto es, el programa devuelve automáticamente el texto etiquetado y un equipo de lingüistas revisa manualmente posibles errores de anotación.

Por lo que se refiere al primer nivel, o anotación morfosintáctica, la parte portuguesa del corpus está siendo anotada con el propio programa eDictor, que incluye un etiquetador morfológico. Para los textos en español, se está utilizando la herramienta FreeLing 3.0 (Padró y Stalinovsky, 2012). A continuación, se muestra un ejemplo de cada tipo de anotación morfológica:

Isto/DEM sei-o/VB-P+CL eu/PRO hoje/ADV ,/, com/P toda/Q-F a/D-F certeza/N ,/, porque/C o/D que/WPRO há/HV-P são/SR-P inimigos/N-P e/CONJ aqueles/D-P home/N do/P+D Regimento/NPR ,/, bem/ADV entendido/VB-PP ,/, soldados/N-P e/CONJ alguns/Q-P oficiais/N-P inferiores/ADJ-G-P ,/, que/WPRO tu/PRO mesmo/FP os/CL conheces/VB-P ,/, estes/D-P são/SR-P aqueles/D-P que/WPRO te/CL fizeram/VB-D o/D mal/N ,/, mas/CONJ Deus/NPR sempre/ADV acode/VB-P a/D-F inocência/N ./.

Fragmento de un texto en portugués anotado con eDictor
CARDS0020, carta familiar de 1827

En_en/SPS00 nuestro_nuestro/DP1MSP país_país/NCMS000 ha_haber/VAIP3S0 habido_haber/VMP00SM muchas_mucho/DIOFP0 tempestades_tempestad/NCFP000 ;_/Fx ha_haber/VAIP30S quitado_quitar/VMP00SM mucho_mucho/DIOMS0 fruto_fruto/NCMS000 ;_/Fx estamos_estar/VAIP1P0 trabajosos_trabajoso/AQ0MP0 ._/Fp Luego_luego/RG ,_/Fc muchos_mucho/DIOMP0 soldados_soldado/NCMP000 ;_/Fx un_uno/DIOMS0 regimiento_regimiento/NCMS000 se_se/P00CN000 va_ir/VMIP3S0 ,_/Fc otro_otro/PIOMS000 entra_entrar/VMIP3S0 ;_/Fx

Fragmento de un texto en español anotado con FreeLing 3.0
PS6001, carta familiar de 1736

Como se puede ver, eDictor asocia a cada palabra del corpus una etiqueta representativa de la clase de palabras a la que pertenece: p. ej., *hoje*, ADV(erbio). Por su parte, el analizador de FreeLing permite obtener, además de la categoría gramatical, el lema de cada palabra: p. ej., *tempestades*, lema *tempestad*,

N(ombre) C(omún) F(emenino) P(lural). En el caso de eDictor, el código de etiquetas está basado en el sistema de anotación manual utilizado por los *Penn Corpora of Historical English* (Kroch, Santorini y Diertani, 2010), ligeramente revisado para adecuarse a las características de la gramática portuguesa⁷. En cuanto al analizador de FreeLing, el etiquetario se basa en la propuesta del grupo EAGLES para la anotación morfosintáctica de lexicones y corpus para todas las lenguas europeas⁸.

Por lo que se refiere a la anotación sintáctica, la metodología adoptada en Post Scriptum varía de nuevo en función de la lengua que es objeto de análisis. Los textos escritos en portugués son etiquetados mediante el sistema de anotación Penn-Helsinki, siguiendo así un estándar de anotación compartido por otros proyectos de sintaxis diacrónica del portugués y desarrollados en cooperación con Post Scriptum: WochWel (URL: <<http://alfclul.clul.ul.pt/wochwel/index.html>>) y Tycho Brahe (Galves y Faria, 2010). Por otro parte, para el conjunto de cartas en español se ha optado de nuevo por la herramienta FreeLing 3.0, que permite tanto un análisis superficial del texto, es decir, la segmentación en constituyentes y la identificación de categorías sintácticas (sintagma nominal, sintagma verbal, sintagma preposicional...), como un análisis de dependencias, más detallado y con información sobre funciones sintácticas (sujeto, objeto directo, objeto preposicional...)⁹.

La etiquetación por separado y mediante diferentes parsers del corpus español y el portugués responde a cuestiones prácticas. Es la única solución que permite hacer uso de la mejor tecnología inmediatamente disponible para el apoyo a los estudios de lingüística diacrónica en estas dos lenguas, al tiempo que permite la cooperación con proyectos paralelos ya en curso: en las Universidades de Lisboa y Campinas para el caso del portugués (proyectos WOChWEL y Tycho Brahe arriba mencionados), y en diversas universidades catalanas para el caso del español (Sánchez-Marco *et al.*, 2010). Somos conscientes de que una decisión como esta dificulta la obtención y el análisis de datos comparativos entre ambas lenguas; no obstante, el sistema de etiquetas utilizado por el etiquetador POS de FreeLing resulta más completo y detallado que el utilizado por eDictor, por lo que siempre es posible convertir automáticamente las etiquetas del primer formato al segundo y anotar sintácticamente una muestra del corpus español siguiendo el sistema de anotación Penn-Helsinki.

5. RELACIÓN ENTRE NIVELES

En los apartados anteriores, se ha ofrecido una visión general sobre el proyecto Post Scriptum, incidiendo especialmente en el proceso de edición de las cartas y su posterior anotación lingüística. A la luz de lo expuesto, se constata

⁷ En <<http://www.tycho.iel.unicamp.br/~tycho/corpus/manual/tags.html>> se puede acceder a la lista completa de etiquetas utilizadas por el analizador morfológico de eDictor [comprobado el 13/09/2013].

⁸ En <<http://nlp.lsi.upc.edu/freeling/doc/tagsets/tagset-es.html#verbos>> se puede acceder a la lista completa de etiquetas EAGLES utilizadas por el analizador morfológico de FreeLing [comprobado el 13/09/2013].

⁹ En <<http://devel.cpl.upc.edu/freeling/svn/trunk/doc/grammars/esCHUNKtags>> se puede acceder a la lista completa de etiquetas utilizadas por el analizador sintáctico de FreeLing [comprobado el 13/09/2013].

que la metodología adoptada en este proyecto de investigación implica diversas tareas o niveles de trabajo que siguen un orden secuencial: búsqueda de cartas en archivos históricos, transcripción del manuscrito en lenguaje XML, normalización ortográfica del texto y anotación lingüística. Con todo, es importante matizar que este conjunto de pasos conlleva, más que un proceso ascendente (de la transcripción semipaleográfica a la anotación sintáctica), una relación dinámica entre diferentes niveles de actuación. De hecho, se da el caso de que algunas normas o decisiones que son tomadas en un determinado nivel obligan a precisar o reformular otras decisiones tomadas en un nivel anterior. Así sucede, por ejemplo, con el tratamiento de las abreviaturas, con las contracciones no estandarizadas o con la puntuación de los textos.

La presencia de abreviaturas, que son de uso frecuente en la escritura cotidiana epistolar, representan un problema para los anotadores automáticos, no sólo por la incapacidad para reconocer la forma correspondiente sino porque una misma abreviatura puede referirse a más de una palabra (por ejemplo, *no.* puede referirse a *nuestro* o a *número*; *pa.* puede referirse a *para* o a *padre*). Para solventar este inconveniente, todas las fórmulas abreviadas aparecen ya expandidas en la transcripción XML mediante el uso de la etiqueta <expan>. De esta forma, la abreviatura se conserva como tal en la edición semipaleográfica, pues el contenido ausente en el original aparece entre paréntesis (*n[úmer]o, pa[dre]*), pero se ofrece ya expandida y sin ninguna marca adicional en la edición modernizada (*número, padre*), precisamente para evitar errores al aplicar el anotador automático.

Respecto a las contracciones no estandarizadas (español: *la venta destas caxas, en lamable compania, ya tescrito*; portugués: *deu pedir, que lle poço dever, pelamor de Deus*), que también aparecen con bastante frecuencia en las cartas, estas pasan a constituir palabras gráficas diferentes en la edición normalizada del texto (español: *la venta d' estas caxas, en l' amable compañía, ya t' he escrito*; portugués: *d' eu pedir, qu' eu lle posso dever, pel' amor de Deus*). Sin embargo, casos como estos impiden mantener el mismo número de tokens en las dos ediciones digitales de la carta, semipaleográfica y modernizada, algo que es necesario respetar para poder sobreponer ambas ediciones en la web y que el usuario pueda cotejar, palabra por palabra, las dos versiones del texto. Por eso, se ha optado por dividir las contracciones no estandarizadas en la edición semipaleográfica (español: *la venta d estas caxas, en l amable compania, ya t e escrito*; portugués: *d eu pedir, qu eu lle poço dever, pel amor de Deus*), de tal forma que se conserve la alineación de tokens con respecto a la edición modernizada. Una etiqueta XML (<note n = "contraction">) permite dejar constancia en la edición semipaleográfica de la existencia de la contracción (*la venta <note n = "contraction">d</note> estas caxas*).

Finalmente, un tercer aspecto cuyo tratamiento está condicionado por tareas posteriores es el de la puntuación de los textos. Ya se ha comentado que la corrección de los signos de puntuación se realiza en el proceso de normalización ortográfica y que es esta versión ya normalizada la que se utiliza como base para el etiquetado automático. En este sentido, la puntuación que presente el texto de entrada es relevante, puesto que puede facilitar el análisis sintáctico de los datos, que es el más complejo y el que demanda una mayor revisión por parte del investigador. Por eso, se ha optado por aplicar una puntuación que,

sin desatender la norma ortográfica, permita aumentar el porcentaje de acierto de los parsers (p. ej., tendencia a frases cortas)¹⁰.

6. OBSERVACIONES FINALES

El proyecto Post Scriptum comenzó en el año 2012 y recibe financiación del Consejo Europeo de Investigación hasta el año 2017. El trabajo realizado hasta la fecha se ha centrado fundamentalmente en la búsqueda de cartas y en la transcripción XML, si bien una parte del corpus portugués cuenta ya con anotación morfológica revisada. La publicación de las cartas se está llevando a cabo en formato electrónico e incluye la siguiente información¹¹:

- Visualización del facsímile del manuscrito.
- Transcripción del texto en XML (edición semipaleográfica).
- Edición normalizada del texto (en la lengua original y en inglés).
- Alineación de edición conservadora y normalizada.
- Anotación morfológica y sintáctica.
- Fichas biográficas de los participantes (autores y destinatarios).
- Palabras clave (áreas de lingüística y de historia).
- Contextualización.

Actualmente, cualquier usuario tiene acceso libre a todos los ficheros XML, a la hoja de estilo y a la DTD; además, en el futuro se podrán consultar los corpus anotados y las extracciones obtenidas a partir de esos corpus. Con todo ello, creemos que el trabajo realizado en Post Scriptum se revelará útil para un público amplio, desde usuarios no especializados pero interesados en episodios históricos de las cartas, hasta historiadores interesados en el tipo de fuentes encontradas o lingüistas interesados en realizar búsquedas sistemáticas en un corpus de textos históricos, casi espontáneos, anotados y no fragmentados. En definitiva, lo que se ofrece es un conjunto de datos que facilita la investigación en línea para múltiples campos de estudio, con diferentes herramientas electrónicas.

BIBLIOGRAFÍA

- Faria, Pablo; Kepler, Fabio y de Sousa, Maria Clara, "An Integrated Tool for Annotating Historical Corpora", *Proceedings of the Fourth Linguistic Annotation Workshop*, 2010, pp. 217-221.
- Galves, Charlotte y Faria, Pablo, *Tycho Brahe Parsed Corpus of Historical Portuguese*, 2010, URL: <<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html>> [29/09/2013].
- Hendrickx, Iris y Marquilha, Rita, "From old texts to modern spellings: an experiment in automatic normalisation", *Journal for Language Technology and Computational Linguistics (JLCL)*, 26(2) (2011), pp. 65-76.
- Kroch, Anthony; Santorini, Beatrice y Diertani, Ariel, *The Penn-Helsinki Parsed Corpus of Modern British English (PPCMBE)*, Department of Linguistics, University of Pennsylvania, CD-ROM, first edition. URL: <<http://www.ling.upenn.edu/hist-corpora/>> [29/09/2013].

¹⁰ Mediante los facsímiles de las cartas, siempre es posible consultar los signos de puntuación originales.

¹¹ La dirección electrónica de Post Scriptum es <<http://ps.clul.ul.pt/index.php>>. En el momento de redactar estas líneas, ya se han publicado en línea 757 cartas.

Nevalainen, Terttu y Tanskanen, Sanna-Kaisa (eds.), *Letter Writing*, Amsterdam, John Benjamins, 2007.

Padró, Lluís y Stanilovsky, Evgeny, "FreeLing 3.0: Towards Wider Multilinguality", *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Estambul, mayo 2012, pp. 2473-2479.

Sánchez-Marco, Cristina; Boleda, Gemma; Fontana, Josep Maria y Domingo, Judith, "Annotation and Representation of a Diachronic Corpus of Spanish", *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Malta, mayo 2010, pp. 2713-2718.



RESUMEN

El proyecto de investigación *Post Scriptum: Archivo Digital de Escritura Cotidiana (P.S.)* tiene por objeto la recuperación y publicación de cartas privadas escritas en España y Portugal durante la Época Moderna. Estas epístolas, en su mayoría inéditas, fueron producidas por autores muy diversos y de diferente condición social. Así, podemos encontrar hombres o mujeres, adultos o niños, amos o criados, ladrones, soldados, artesanos, curas, políticos, y otro tipo de categorías sociales. Sus misivas sobrevivieron excepcionalmente, al cruzarse sus vidas con los medios de persecución tanto de la Inquisición como de distintos tribunales civiles y eclesiásticos, instituciones que solían hacer uso de la correspondencia privada como prueba de los delitos que estaban siendo juzgados. Estas fuentes escritas suelen presentar una retórica (casi) oral, tematizando asuntos cotidianos que hasta ahora no se habían estudiado más que a partir de casos aislados. En este trabajo, se explicará la metodología utilizada en la edición digital de los documentos para su disponibilidad en línea; además, se dará a conocer el proceso de modernización de los textos y su posterior etiquetación morfológica y sintáctica. El objetivo final es elaborar un corpus diacrónico anotado que sirva como recurso electrónico para el estudio lingüístico e histórico del español y el portugués.

Palabras clave: Cartas, portugués, español, anotación de corpus, lingüística diacrónica.

ABSTRACT

Post Scriptum: A Digital Archive of Ordinary Writings (P.S.) is a project that aims to collect and publish Portuguese and Spanish private letters written along the Modern Ages. These documents are unpublished epistolary writings, written by authors from different social backgrounds. They could be either masters or servants, adults or children, men or women, thieves, soldiers, artisans, priests, political activists, among other kinds of social agents. Their epistolarity survived by chance, when their paths met the persecution means used by the Inquisition and the civil courts, two institutions that used private correspondence as criminal evidence. These textual resources often present an (almost) oral rhetoric, treating everyday issues of past centuries in a register that hasn't been easy to study, apart from brief examples. In the proposed paper, discussion over the methodological options that lead to the digital edition available online will be raised. Further, the modernization of texts the POS and syntactic annotation will be explained. The aim is to develop a diachronic and annotated corpus that could be used as an electronical resource for linguistic and historical research of Spanish and Portuguese.

Keywords: Letters, Portuguese, Spanish, corpus annotation, diachronic linguistics.



La revista digital
Janus. Estudios sobre el Siglo de Oro
está patrocinada por el
SIELAE



SIELAE

www.bidiso.es/sielae/



UNIVERSIDADE DA CORUÑA