



La irrupción de las humanidades digitales en España y su aplicación filológica¹

Enrique Ripoll
Universidad de Alicante

RESUMEN:

En este artículo se hace un recorrido histórico de la influencia que han tenido los ordenadores en las investigaciones humanísticas y, en concreto, su aparición y evolución en España en el último cuarto del siglo XX. Comienza en los años 40, momento en el que aparecieron las primeras computadoras electrónicas, como breve introducción, para posteriormente, detallar los momentos clave que supusieron un paso evolutivo impulsado por lingüistas que supieron aprovechar su potencial. Aunque en España la adopción de los ordenadores para la investigación fue más tardía que en otros países, este hecho no impidió que rápidamente se creasen investigaciones interdisciplinarias que unían la informática y la filología.

PALABRAS CLAVE: humanidades digitales hispánicas, lingüística computacional, procesamiento de datos, historia informática, informática textual.

ABSTRACT:

This article provides a historical review of the influence that computers have had on humanistic research and, specifically, its appearance and evolution in Spain in the last quarter of the 20th century. It begins in the 40s, when the first electronic computers appeared, as a brief introduction, to later detail the key moments that marked an evolutionary step driven by linguists who knew how to take advantage of their potential. Although in Spain, the adoption of computers for research was later than in other countries, this fact did not prevent the rapid creation of interdisciplinary investigations that linked computer science and philology.

KEYWORDS: Hispanic digital humanities, computational linguistics, data processing, computer history, textual computing.

La informática nació unida a las matemáticas. Se concibió originalmente como un mecanismo para acelerar los cálculos repetitivos que se venían haciendo manualmente; sin embargo, tan pronto como se extendió su uso a finales de los años 40, ciertos humanistas se percataron de que, si hacían uso de un ordenador, podían conseguir resultados que, de otra forma, eran inabordable por tiempo o recursos. Las humanidades digitales comenzaban su andadura.

1.- Esta publicación es parte del proyecto de I+D+I «Poesía, ecdótica e imprenta» (PID2021-123699NB-I00), financiado por MCIN/AEI/10.13039/501100011033/ y «FEDER Una manera de hacer Europa».

Para que una máquina se considere un ordenador o computadora tiene que cumplir una serie de requisitos y no solo limitarse a hacer cálculos u operaciones matemáticas. En ese caso sería tan solo una calculadora, que, dada una *entrada*, efectúa una serie de operaciones y entrega una *salida*. Por supuesto, debe ser capaz de hacer eso, pero, además, tiene que ser *universal*, concepto que debemos a Alan Turing, quien definió, mediante un modelo matemático, un dispositivo teórico conocido como *Máquina de Turing* (1937: 232-233). Describió con exactitud su funcionamiento y características, entre las cuales destacaba, precisamente, su *universalidad*. Hasta ese momento, los dispositivos de cálculo que se construían se diseñaban específicamente para que efectuasen una única función, como, por ejemplo, resolver ecuaciones diferenciales (Ralston & Reilly 1983: 533). Pero Turing fue más allá e imaginó una máquina a la que se le podía indicar lo que debía hacer con las entradas que recibía o, dicho de otra forma, se podía programar. Con ello, sentaba las bases de lo que se consideraba un ordenador.²

El modelo teórico de Turing se llevó rápidamente a la práctica. La inestabilidad geopolítica mundial de los años 40 favoreció que las grandes potencias invirtieran en el diseño y construcción de computadoras. En el campo militar, demostraron ser capaces de realizar tareas tan diversas como descifrar mensajes codificados por el bando enemigo o calcular trayectorias de misiles. De los primeros diseños que se debían programar mediante cables se evolucionó rápidamente a máquinas que poseían una memoria en la que residían las instrucciones. Pero, finalizada la contienda, había que darles otros usos y, en Estados Unidos, dos de los principales ingenieros que habían trabajado en la construcción de computadoras decidieron crear una empresa para venderlos a gran escala. En 1951 veía la luz el primer ordenador comercial, que no estaba, por tanto, expresamente destinado a fines militares: el UNIVAC I, seguido dos años después por el IBM 701 (Ralston & Reilly 1983: 535-537).

Aunque estos ordenadores se habían construido expresamente con la idea de efectuar operaciones matemáticas y analizar datos a gran escala en las empresas e instituciones, esta idea preconcebida no impidió que se les encontrasen otros usos en diferentes campos y, en concreto, en el que nos ocupa: la filología. De hecho, tan pronto como aparecieron públicamente, se creó el primer gran proyecto de humanidades digitales, que comenzó en 1950 y se dio por finalizado en 1980: el *Index Thomisticus*, con el apoyo de IBM y liderado por el padre Busa (1980: 87), cuyo objeto último era la obtención de las concordancias de la obra de Santo Tomás de Aquino.

La recopilación y análisis de textos no fue la única aplicación lingüística de la informática: el otro campo en el que se vislumbró su potencial fue el de la traducción. En este caso, sus impulsores fueron Warren Weaver y Donald Booth quienes, ya a finales de los 40, debatieron sobre las posibilidades de los computadores en la traducción automática (Booth & Locke 1955: 2) y, en 1952, se organizó el primer congreso sobre esta cuestión:

Eighteen scholars, including Booth as the only non-American delegate, gathered on the first international conference on Machine Translation at MIT, followed by a meeting later that year in London where some forty linguists met during the

2.- El concepto de máquina programable se lo debemos a Babbage y su *Máquina Analítica*. Aunque es un dispositivo mecánico que no se llegó a construir, Ada Lovelace diseñó un algoritmo que calculaba los números de Bernoulli con el que demostró su capacidad de programarse (Menabrea 1843: 722-731).

International Linguistic Congress. A year later, Machine Translation appeared for the first time in a scholarly textbook written by Andrew and Kathleen Booth. In their book *Automatic Digital Calculators*, aimed at a readership of computer scientists, the authors published a chapter on *Some applications of computing machines* in which Machine Translation was discussed at length (Vanhoutte 2013: 124).

En España, hubo que esperar hasta 1958 para que llegase el primero ordenador, destinado a la compañía ferroviaria RENFE, un IBM 650;³ poco tiempo después, lo hizo un UNIVAC UCT, esta vez para la Junta de Energía Nuclear (Arroyo Galán 2005: 49). Sin embargo, pese a estas rápidas adquisiciones iniciales, no sería hasta diez años más tarde, en 1968, cuando se pondría en funcionamiento la primera máquina dedicada a la investigación en el —expresamente construido para tal fin— Centro de Cálculo de la actual Universidad Complutense de Madrid: un IBM 4090. Se trataba de un computador cedido por el fabricante en cuestión, que llegaba a ocupar toda una planta y que ya había estado operativo previamente en el CERN de Ginebra y en la Universidad de Heidelberg (López & Munarriz 2021: 22-30).

El Centro de Cálculo se consideraba una instalación singular, dado que, ya desde sus inicios, se incentivó el uso de la nueva máquina más allá de los cálculos científicos. Se organizaron seminarios para reunir humanistas con el objetivo de que aportasen ideas innovadoras, como el de Generación Automática de Formas Plásticas, que comenzó a celebrarse durante el curso 1968-1969 con una periodicidad anual (Castaños Alés 2000: 86). Además, se crearon becas en «un empeño en que artistas, arquitectos o músicos aprendieran un lenguaje de programación» (López & Munarriz 2021: 28).

Fue, precisamente, en estas instalaciones donde comenzó su carrera uno de los primeros investigadores que se dedicó a las humanidades digitales: Francisco Marcos Marín. En un artículo fechado en 1971, describía un sistema de traducción automática inglés-francés, que llevaba cinco años desarrollándose en la Universidad de Montreal, y las implicaciones que tendría su adaptación al español (Marcos Marín 1971: 313). Era el mismo año en el que la Real Academia Española (RAE) decidía investigar sobre las posibilidades de aplicación de los ordenadores a sus diccionarios contratando a dos investigadores: Ignacio Soldevila Durante, para estudiar la posible informatización del diccionario académico, y el propio Francisco Marcos Marín, para colaborar en el *Diccionario Histórico* (2009: 392).⁴ Sin embargo, la RAE no disponía de equipamiento informático y, por tanto, también fue el Centro de Cálculo de la Universidad Complutense de Madrid el que acogió, *de facto*, los primeros proyectos de humanidades digitales en España. Allí fue donde se desarrolló el primer analizador sintáctico automático de español (Campo *et al.* 1973: 16) y el primer atlas lingüístico plurilingüe que hacía uso de un ordenador para su representación gráfica (Ariza *et al.* 1973: 12).

La producción científica de Marcos Marín es cuantiosa y su revisión completa supera los objetivos de este trabajo, pero cabe destacar su liderazgo como director científico del equipo que creó *Admyte*, un archivo digital de manuscritos y textos españoles, que incor-

3.- Actualmente se encuentra expuesto en el Museo Nacional de Ciencia y Tecnología de A Coruña.

4.- El objetivo inicial era que Marcos Marín colaborase en los arabismos del *Diccionario Histórico*; sin embargo, su director, Rafael Lapesa, decidió enviarlo a la Escuela de Pisa, que en aquel momento era un referente en lingüística computacional (2009: 392).

poraba un buscador que permitía la consulta de las bases de datos, de las transcripciones y de los facsímiles digitales (Marcos Marín 1994: 196-217). Fue editado en tres CD-ROMs y, actualmente, aún es accesible online bajo suscripción:⁵

El trabajo en equipo se puso de manifiesto en *Admyte 0*, disco que contenía un volumen ingente de información de diversa naturaleza: 64 textos medievales transcritos por diversos investigadores en el marco que les brindaba *el Seminary of Medieval Spanish Studies* de la Universidad de Wisconsin-Madison; los catálogos generales de manuscritos e incunables en español (BETA/BOOST), catalán (BITECA/BOOCT) y portugués (BITAP/BOOPT); un programa de recuperación y análisis textual para los estudios lingüísticos y literarios (TACT); y un programa para preparar ediciones críticas capaz de cotejar hasta 30 versiones diferentes de la misma obra (UNITE).

Los materiales de *Admyte I* fueron seleccionados y elaborados *ex professo* por el equipo editorial, que trabajó en estrecha relación con la Biblioteca Nacional en el marco de la celebración del V Centenario del Descubrimiento de América. En total, se recogieron las transcripciones y las imágenes digitalizadas de 61 obras transmitidas en libros incunables y postincunables, que dan una idea muy precisa del panorama cultural español y europeo en los años del Descubrimiento de América. Respecto de estos dos discos previos, *Admyte II* presenta dos cambios de importancia: sólo ofrece las transcripciones de las obras sin imágenes digitalizadas que las acompañen; a cambio, el número de los textos ha crecido notablemente, pues se han incorporado nada menos que 165 nuevas ediciones de obras correspondientes en su mayor parte a la España medieval, aunque tampoco falte un importante testimonio renacentista, el Lazarillo de Tormes en dos de sus testigos principales, y hasta una joya del barroco literario español, el manuscrito Chacón con la obra poética de don Luis de Góngora. Además, *Admyte* hace justicia al romancero por partida doble, pues remedia su ausencia de nuestro catálogo de obras medievales (BETA/BOOST) y corrige el injustificable olvido de este fascinante corpus en los dos discos previos de *Admyte*.⁶

Otro de los pioneros españoles en las humanidades digitales fue el lexicógrafo Manuel Alvar Ezquerro, coetáneo de Marcos Marín, quien, a mediados de los años 70, tuvo el privilegio de trabajar junto a Bernard Quemada⁷ en el *Trésor de la langue française*, un diccionario histórico en el que se utilizaron medios informáticos para su creación (Marcos Marín 2009: 393). Su primer artículo sobre humanidades digitales se publicó a mediados de los 70 y trata la obtención de índices de rimas y sufijos utilizando una computadora, con el fin de evitar el trabajo manual y, además, acelerando con ello el proceso (Alvar Ezquerro 1976: 35-36). Pero sus innovaciones metodológicas no quedaron relegadas a la investigación, sino que, además, las llevó a las aulas desde muy temprano, al hacer uso de

5.- Entre el equipo de *Admyte* estaba Charles B. Faulhaber, que, como se comentará más adelante, será el principal responsable de crear la web *PhiloBiblon* (Faulhaber 2014: 17), un metacatálogo online de bibliografías ibéricas.

6.- La cita está extraída de la sección *Historia* de la web *Admyte*, <<https://www.admyte.com/admyteonline/historia.htm>> [consulta: 02/04/2023].

7.- Bernard Quemada fue un lexicógrafo nacido en San Sebastián, de padre español y madre francesa, que se formó y trabajó en Francia. Estuvo al frente de «l'elaboració del *Trésor de la Langue Française*, un dels primers corpus textuais digitalitzats que es van crear a Europa durant la segona meitat del segle XX» (Cabré Castellví 2019: 714-715).

la automatización de los estudios lingüísticos en sus cursos de Lingüística Aplicada en la Universidad Complutense de Madrid (Pêcheux 1978: 7).

Una década después de la instalación del primer computador para uso científico en España, ya era evidente que los ordenadores no solo servían para hacer cálculos, sino que se podían utilizar en las investigaciones humanísticas. Las aportaciones en el campo de la lingüística computacional eran cada vez más numerosas y solo desde estas circunstancias se explica que, en 1983, Montserrat Meya y María Felisa Verdejo creasen la revista científica *Procesamiento de Lenguaje Natural* en la Facultad de Informática de la Universidad del País Vasco, con unas líneas de trabajo abiertas a la colaboración entre distintas ramas de conocimiento, tal como expresaban en su presentación:

El propósito que perseguimos con ella es, ante todo, crear un órgano especializado de comunicación a través del cual podamos conocer de forma directa los trabajos, investigaciones y proyectos que se realizan en España dentro del área del Proceso del Lenguaje Natural. Por ello, al tratarse de una tarea interdisciplinaria, nos hemos reunido en un mismo empeño profesionales de diferente formación: ingenieros, informáticos y lingüistas (Meya 1983: 1).

Esta interdisciplinaridad ya venía reflejada en la formación de sus editoras, dado que Meya⁸ es lingüista y Verdejo, informática. Meya trabajaba para Siemens, mientras que Verdejo, al igual que Marcos Marín, se inició y ejerció en las humanidades digitales en el marco del Centro de Cálculo de la Universidad Complutense de Madrid:

El día 30 de octubre de 1975, en la Universidad de París VI (Institut de Programmation) leyó su tesis doctoral de Tercer Ciclo, sobre el tema «Un estudio del lenguaje natural y su aplicación a un diálogo en castellano con un robot» la Analista del Centro de Cálculo de la Universidad Complutense de Madrid Srta. María Felisa Verdejo; la tesis fue juzgada con la calificación de «Très honorable» por un tribunal presidido por el profesor Arzac y entre los miembros del mismo figuraba D. Ernesto García Camarero Director del Centro de Cálculo de la Universidad Complutense (García Camarero 1976: 102).

El primer artículo de Verdejo (1976: 3)⁹ sobre humanidades digitales parte de su tesis doctoral: un sistema de interacción de pregunta-respuesta en español con una máquina, dotado de cierta inteligencia en sus respuestas, en el cual no solo aplica el procesamiento de lenguaje natural, sino también técnicas de inteligencia artificial, los dos campos en los que Verdejo ha basado mayormente su investigación. Esta dualidad la llevó a fundar, en 1983, la *Sociedad Española para el Procesamiento del Lenguaje Natural* (SEPLN) junto a la revista antes mencionada, así como la *Asociación Española para la Inteligencia Artificial* (AEPIA). Como reconocimiento a su trayectoria, en 2014 se le concedió el Premio Na-

8.– Montserrat Meya recibió el premio Ada Byron a la Mujer Tecnóloga en el 2014, un galardón creado por la Facultad de Ingeniería de la Universidad de Deusto y que puede consultarse en <<https://www.deusto.es/es/inicio/somos-deusto/facultades/ingenieria/eventos-y-premios/premio-ada-byron>> [consulta: 27/08/2023].

9.– El primer artículo de Verdejo, en realidad, es anterior y en colaboración, sobre la creación de un preprocesador para un lenguaje de programación, el ALGOL 60 (Garijo & Verdejo 1973). Un *preprocesador* es una aplicación que, dada una entrada en un lenguaje formal, lleva a cabo un análisis léxico y sintáctico y produce una salida en otro lenguaje que permite su traducción directa, con el uso de un compilador, a código interpretable por el ordenador.

cional de Informática José García Santesmases y en 2016 fue nombrada *Doctora Honoris Causa* por la Universidad de Alicante.¹⁰

Investigadores como Marcos Marín, Alvar, Meya y Verdejo, entre otros muchos, impulsaron el campo lingüístico de las humanidades digitales en los años 70, que se consolidó y se expandió a otras universidades españolas en los 80. Buena muestra de ello es el hecho de que el primer número de *Procesamiento de lenguaje natural* esté formado por artículos de investigadores de las universidades del País Vasco, Castilla-La Mancha,¹¹ Valencia, Barcelona, Granada y Complutense de Madrid, además de otros dos que trabajaban en las empresas privadas IBM y Siemens.¹²

Fue una década marcada por una actitud aperturista que propició la colaboración en proyectos internacionales como *EUROTRA*, el traductor multilingüe europeo. Había un interés creciente en este tipo de herramientas, por lo que, no solo hubo inversión pública en la traducción automática del español, sino que las grandes empresas informáticas también crearon sus propios proyectos, como IBM con *MENTOR* o SIEMENS con *METAL* (Marcos Marín 2009: 394-395).

Pero, pese a este interés creciente en la aplicación de la computación en el campo lingüístico, su incorporación en la vertiente literaria de la filología fue más tardía:

A finales de los años 80 se conocían algunas iniciativas de filólogos que empleaban herramientas computacionales, pero eran rarísimas excepciones, y desde luego ninguna tuvo impulso institucional ni ayudas económicas antes de los años 90. Fue en 1990 cuando el Gobierno Español, con motivo de las Conmemoración del V Centenario del Descubrimiento de América, creó una *Comisión Nacional* y se nombró director del área de *Industrias de la Lengua de la Sociedad Estatal para la Ejecución de los Programas del Quinto Centenario (1990-1992)* a Francisco Marcos Marín, profesor de la Universidad Autónoma de Madrid, que había estado en contacto con grupos norteamericanos en uno de los proyectos primitivos más ambiciosos que hoy conocemos como *Philobiblon*, base de datos biobibliográfica sobre textos romances escritos en la Península Ibérica en la Edad Media y el Renacimiento (López Poza 2020: 133).

En la década de los 90 es cuando hizo su aparición la *World Wide Web* y, con ella, la universalización de Internet. Todo el mundo quería estar en ese escaparate mundial que se estaba extendiendo rápidamente más allá del ámbito investigador y, por supuesto, los grandes proyectos humanísticos no fueron ajenos a su influencia.¹³ Tal fue el caso de *Phi-*

10.– Según atestigua el currículum que se elaboró para su nombramiento como *Doctora Honoris Causa* por la Universidad de Alicante y que se puede consultar en <<https://web.ua.es/es/protocolo/documentos/eventos/honoris/verdejo-maillio-felisa-2016/apunte-biografico-de-felisa-verdejo-maillio.pdf>> [consulta: 02/04/2023].

11.– En aquel momento era el Colegio Universitario de Ciudad Real, que después se ha convertido en la Universidad de Castilla-La Mancha.

12.– Los artículos están firmados en este orden por los siguientes investigadores: F. Ares de Blas (Facultad de informática, San Sebastián), N. Antonio Campos (Colegio Universitario, Ciudad Real), F. Casacuberta y E. Vidal (Centro de Informática, Valencia), M. Meya (Siemens, Munich), J. Rafel (Dpto. Catalán. Universidad de Barcelona), J. Rubio Ayuso y M.C. Carrión Pérez (Universidad de Granada), Martín S. Ruipérez (Facultad de Filología, Universidad Complutense), L. de Sopena (Centro Científico IBM, Madrid), M.F. Verdejo (Facultad de Informática, San Sebastián).

13.– José Manuel Lucía (2010) destaca, en este sentido, tres de ellos en cuanto a la filología hispánica que apostaron por la publicación en línea como el futuro: *LEMIR*, que comenzó a gestarse entre 1994 y 1995; *PhiloBiblon*, en 1997; y la *Biblioteca Virtual Miguel de Cervantes*, que tuvo su primera versión en julio de 1999.

loBiblon que, después de su incorporación en *Admyte*, publicaba su versión *online* en 1997. Aunque actualmente está formado por un compendio de diferentes catálogos bibliográficos de textos ibéricos, sus inicios fueron en la década de los 70 y se limitaba a la *Bibliography of Old Spanish Texts* (BOOST):

Se concibió en un primer momento como un paso inicial necesario para el empleo del ordenador en la confección del Diccionario de español medieval (DOSL) en el que la Universidad de Wisconsin, Madison, viene trabajando desde hace aproximadamente medio siglo. Este ambicioso proyecto lexicográfico tiene como primer propósito la formación y publicación de un amplio corpus léxico por medio de citas, reflejando el empleo de cada término tal como lo presentan los diferentes documentos —impresos o manuscritos— producidos antes del año 1501 (BOOST 1984: XVII).

Estos materiales dieron lugar a la *Bibliografía Española de Textos Antiguos* (BETA), que se recogieron en 1993 en un CD-ROM, como parte de *Admyte*. De igual manera, Beatrice Concheff desarrolló en 1985 la *Bibliography of Old Catalan Texts* (BOOCT), que acabó dando lugar a la *Bibliografía de Textos Antics Catalans, Valencians i Balears* (BITECA).¹⁴ Ambas bibliografías completaron la perspectiva ibérica con la *Bibliografía de Textos Antigos Galegos e Portugueses* (BITAGAP), derivada, en última instancia, de la *Bibliography of Old Portuguese Texts* (BITAP/BOOPT).¹⁵ BITECA y BITAP también formaron parte del CD-ROM de *Admyte* antes de llegar a estar disponibles en la red, en este último caso ya como BITAGAP y a diferencia de la *Bibliografía de Poesía Áurea* (BIPA), que se ha sumado a este portal ya avanzado el siglo XXI.¹⁶ BIPA había permanecido hasta entonces *offline* y el acceso se llevaba a cabo mediante consultas personales que los investigadores hacían a sus creadores, José J. Labrador y Ralph DiFranco. Aunque la base académica de *Philobiblon* se encuentra en la University of California, la mayoría de sus responsables pertenecen a universidades españolas, en las que diferentes investigadores principales han coordinado sucesivos proyectos de investigación de financiación pública, de manera independiente para cada una de estas bibliografías.¹⁷

14.– BITECA, en un principio, era tan solo la *Bibliografía de Textos Antics Catalans*, aunque se corrigió el nombre a partir de un convenio con la Acadèmia Valenciana de la Llengua.

15.– «Deriva de la *Bibliography of Old Spanish Texts* (BOOST), cuya primera edición impresa vio la luz en 1975 (Cárdenas *et al.*, 1975) como repertorio de textos y manuscritos medievales en castellano, una herramienta ancilar del *Dictionary of the Old Spanish Language* de Madison. Con el tiempo BOOST, rebautizada como BETA (*Bibliografía Española de Textos Antiguos*), junto con sus congéneres BITAGAP (*Bibliografía de Textos Antigos Galegos e Portugueses*) y BITECA (*Bibliografía de Textos Antics Catalans, Valencians i Balears*) pasaron al formato CD-ROM como parte de ADMYTE (*Archivo Digital de Manuscritos y Textos Españoles*) en 1993» (Faulhaber 2009: 191).

16.– «La *Bibliografía de la Poesía Áurea* es una base de datos digital de la poesía española de los siglos XVI y XVII recogida en fuentes manuscritas e impresas. La versión actual, parcial e incompleta, se ha montado únicamente para hacer pruebas» (BIPA 2023). Los responsables de BIPA, según su web el 03/02/2023, son: Ralph DiFranco (University of Denver) y José J. Labrador Herraiz (Emérito, Cleveland State University).

17.– El equipo de BETA, según su web el 03/02/2023, está formado por Charles B. Faulhaber (University of California, Berkeley), Angel Gómez Moreno (Universidad Complutense de Madrid), Nicasio Salvador Miguel (Universidad Complutense de Madrid), Antonio Cortijo Ocaña (University of California, Santa Barbara), María Morrás, (Universitat Pompeu Fabra / Oxford University), Óscar Perea Rodríguez (Universidad Complutense de Madrid), Álvaro Bustos Tàuler (Universidad Complutense de Madrid) y José Luis Gonzalo Sánchez-Molero (Universidad Complutense de Madrid). Los responsables de BITECA, según su web el 03/02/2023, son: Gemma Avenoza (+) (Universitat de Barcelona, Institut de Recerca en Cultures Medievales), Lourdes Soriano (Universitat de Barcelona, Institut de Recerca en Cultures Medie-

A pesar de que acabó derivando en la creación de *Philobiblon*, la BOOST se originó, en realidad, en el *Hispanic Seminary of Medieval Studies* (HSMS) de la Universidad de Wisconsin-Madison como un proyecto complementario para crear un diccionario electrónico: el *Dictionary of the Old Spanish Text* (DOSL), cuya versión en papel se venía elaborando desde 1935. En 1971, se estudió la posibilidad de generarlo a partir de un corpus electrónico del español medieval y se llegó a la conclusión de que era necesario crear previamente una bibliografía que recogiese los ejemplares de incunables que existían, con el fin de seleccionar el material para confeccionarlo (Nitti 1978: 43). En 1974 apareció la primera versión de BOOST, a la que le siguieron dos más en papel. Por su parte, la creación del corpus electrónico comenzó en 1979 con la edición, en microficha, de *The Concordances and Texts of the Royal Scriptorium Manuscripts of Alfonso X*.¹⁸ Desde el 2011, la colección se publica en la web del seminario como *Biblioteca digital de textos del español antiguo*.¹⁹ Esta serie sirvió de inspiración al *Seminari de Filologia i Informàtica* de la Universidad Autònoma de Barcelona para crear el *Arxiu Informatitzat de Textos Catalans Medievals*, coordinado por Lola Badia, José M. Bleca, Glòria Claveria, Josep Pujol, Amadeu Soberanas y Joan Torruella (1995). El proyecto, en realidad, comenzó en 1986 «amb vista a confeccionar un fons de textos apte per a una sèrie d'ànalis informàtiques que pretenen acostar-nos amb major precisió a la interpretació i organització de les dades paleogràfiques, textuals i lingüístiques proporcionades pels manuscrits originals» (Torruella & Lawrance 1988-1989: 31). Aunque las primeras microfichas con los textos y las concordancias respectivas generadas a partir del *Oxford Concordance Program* se publicaron entre 1995 y 1996 (Torruella 1991: 245),²⁰ ya se habían aplicado técnicas computacionales sobre este corpus de cancioneros catalanes medievales en 1992, para generar un índice de frecuencias de las rimas del *Cançoner L*, así como un rimario alfabético en el que se mostraba, junto a su autor, la consonancia, frecuencia y palabras-rima utilizadas en esa colección poética (Torruella 1992: 5-9). Este es un ejemplo temprano de la rentabilidad que suponía el tratamiento de corpus desde las humanidades digitales para la obtención de resultados de interés literario y no exclusivamente lingüístico.

vals) y Vicenç Beltran (Universitat de Barcelona, Università di Roma «La Sapienza»). Y, finalmente, los recopiladores de BITAGAP desde 1988, según su web el 03/02/2023, son: Arthur L-F. Askins (University of California, Berkeley), Harvey L. Sharrer (University of California, Santa Barbara), Martha E. Schaffer (Universidade de San Francisco) y Aida Fernanda Dias (+) (Universidade de Coimbra). En asociación con desde el 2008 con: Cristina Sobral (Faculdade de Letras, Universidade de Lisboa), Pedro Pinto (Centro de Estudos Históricas, Universidade Nova de Lisboa), Filipe Alves Moreira (Universidade do Porto), Mariña Arbor Aldea (Facultade de Filoloxía, Universidade de Santiago de Compostela), Maria de Lurdes Rosa (Departamento de História, Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa) y Ricardo Pichel Gotérrez (Universidad de Alcalá).

18.- Se puede consultar la historia completa del HSMS en <<http://www.hispanicseminary.org/history-en.htm>>.

19.- A fecha de la consulta, dispone de 12 corpus de distinta temática que agrupan un total de 382 textos de libre acceso a través de <<http://www.hispanicseminary.org/textconc-es.htm>> [consulta: 14/04/2023].

20.- *Cançoner L* (Barcelona, Biblioteca de Cataluña, ms. 9); *Espill* de Jaume Roig (Ciudad del Vaticano, Biblioteca Apostólica, ms. Latín 4806); *Cançoneret de Ripoll* (ms. 129 del fondo de Ripoll del Archivo de la Corona de Aragón); *Cançoner dels Masdovelles* (Barcelona, Biblioteca de Cataluña, ms. 11) y *Cançoner del Marquès de Barberà* (Biblioteca del Monestir de Montserrat, ms. 992). Se trataba de cinco juegos de microfichas, que, posteriormente, se completaron con otros cinco: el *Cançoner Vega-Aguiló* (Barcelona, Biblioteca de Catalunya, ms. 7 i 8); *Cançoner de París* (París, Biblioteca Nacional, ms. esp. 225); *Cançoner de l'Ateneu* (Barcelona, Biblioteca de l'Ateneu Barcelonès, ms. 1); *Jardinet d'orats* (Barcelona, Biblioteca de la Universitat, ms. 151); y *Cançoner de Saragossa* (Zaragoza, Biblioteca Universitaria, ms. 184).

También en 1995, y con más de 30 años de diferencia respecto al primer corpus electrónico de la lengua inglesa,²¹ la RAE acometía el proyecto de crear dos corpus: el *Corpus de referencia del español actual* (CREA) y el *Corpus diacrónico del español* (CORDES), que, respectivamente, atendían a la perspectiva sincrónica y diacrónica de la lengua, cuya primera versión se ponía en línea en 1998 (Rojo 2016: 200).

Ambos conjuntos son complementarios, de manera que el CREA contiene los textos pertenecientes a los últimos treinta años de historia del español, mientras que el CORDE se ocupa de todo lo demás. El carácter integrado de ambos corpus se refleja en la previsión de que los textos pertenecientes a períodos que, por el paso del tiempo, vayan quedando fuera del ámbito del CREA pasarán a formar parte de CORDE (Sánchez Sánchez & Domínguez Cintas 2007: 137).²²

No obstante, en el diseño del CREA se estableció un porcentaje de textos distinto a cada quinquenio, comenzando por un 10% y añadiendo un 5% adicional en cada salto temporal. De esta manera, los que se situaban entre el 1975 y 1979 suponían un 10%, los comprendidos entre 1980 y 1984, un 15%, y así sucesivamente. Esta característica implicaba que, al traspasarlos al CORDE, estos porcentajes debían reajustarse para que todos tuviesen el mismo peso, lo que habría requerido, en última instancia, la eliminación de textos.²³ Es por ello que, finalmente, se decidió no trasvasar información al CORDE, cerrar la incorporación de nuevos registros en el CREA en 2008 y ceder el testigo a un nuevo corpus: el *Corpus del español del siglo XXI* (CORPES) (Rojo 2016: 200-202).

Los proyectos literarios de esta época se apoyaron en la informática para crear bases de datos y eliminar el papel. La difusión de los resultados de las investigaciones en Internet suponía modificar el modelo tradicional de publicación impresa. Era un cambio de paradigma que, a pesar de que tres décadas después se ha consolidado totalmente en el caso de las revistas científicas, en ese momento suponía toda una revolución y un cambio profundo del proceso editorial. Solo unos visionarios se atrevieron a romper con ese modelo tradicional y dar el salto al mundo digital. Tal fue el caso de José Luis Canet, de la Universitat de València, y Sagarrio López Poza, de la Universidade da Coruña.

Canet se encargó de dar vida a LEMIR a mediados de los 90, pero su relación con la informática había comenzado una década antes, desde la propia aparición de los ordenadores personales. Se comenzó interesando por las bases de datos relacionales, con las

21.– En 1964 se publicaba el primer corpus electrónico del inglés, el *Brown University Standard Corpus of Present-Day American English*, más conocido como *Brown Corpus* (Francis & Kucera 1979). La primera versión estaba basada en los textos de los libros y periódicos publicados durante 1961 (Francis 1965: 267). Respecto a los primeros corpus de español anteriores al CREA y CORDE, Rojo (2015) enumera los siguientes: el *Corpus de Lovaina* bajo la dirección de Josse de Kock; ENTREVIS90 y ENTREVIS95, construidos por Kjær Jensen; el *Corpus Vox-Bibliograf* (CVB), dirigido por Manuel Alvar Ezquerro; el corpus CUMBRE, dirigido por Aquilino Sánchez; el *Corpus del español mexicano contemporáneo* (CE-MC), que ha permitido la creación de varios diccionarios dirigidos por Luis Fernando Lara; los corpus multilingüe CRATER, NERC y PAROLE, surgidos en el marco de proyectos europeos; el *Corpus lingüístico de referencia de la lengua española en Argentina*, el *Corpus lingüístico de referencia de la lengua española en Chile* y el CORLEC, los tres dirigidos por Francisco Marcos Marín; el corpus LEXESP, elaborado por un equipo de lingüistas y psicólogos; el *Corpus of Contemporary Spanish* elaborado por Barry Ife; y ADMITE, confeccionado por Francisco Marcos Marín, Charles Faulhaber, Ángel Gómez Moreno y Antonio Cortijo Ocaña.

22.– Rojo (2015: 200) establece el marco temporal del CREA en 25 años.

23.– Se puede encontrar una explicación más detallada de los porcentajes de cada periodo temporal y la problemática que acarrea el pasar los textos del CREA al CORDE en Rojo 2016.

que creó un gestor bibliográfico, hecho que le valió para que Apple lo incluyese entre sus desarrolladores en 1987 y que, en 1990, fuese nombrado director de las Bibliotecas de la Universitat de València con la misión de modernizarlas tecnológicamente. Fruto de ese trabajo apareció online en 1993 la web del *Servicio de Información Bibliográfica* (SIB) de la Universitat de València, experiencia que aprovechó para lanzar posteriormente el portal *LEMIR*,²⁴ enfocado a la literatura medieval y renacentista, en cuyo marco apostó por una de las primeras revistas de literatura con formato únicamente electrónico: *LEMIR. Revista Española Medieval y del Renacimiento* (Canet 2014: 11-12). Su primer número apareció en 1995-1996 con tres artículos y cinco ediciones de texto. El portal, en su primera versión, además de la revista electrónica, ofrecía noticias y eventos de interés, ediciones de textos de difícil publicación en editoriales comerciales y un par de secciones con entidad propia: *Exemplaria* y *Tirant*, que, posteriormente, se convirtieron en revistas electrónicas. La primera estaba dedicada a la literatura sapiencial y dirigida por Marta Haro,²⁵ mientras que *Tirant*, a cargo de Rafael Beltrán, se centraba en la literatura caballerescas. El primer número de *Exemplaria*, de 1996, recoge un repertorio de novedades bibliográficas, un apartado de textos con dos transcripciones y un último apartado de miscelánea de investigación que da cabida a noticias en general. Por su parte, el primer número de *Tirant* se materializó dos años después con cuatro artículos, seis reseñas, un apartado de bibliografía, otro de novedades bibliográficas y el resumen de una tesis.

El concepto de *LEMIR* fue un éxito, por lo que, en 1998, vio la luz un proyecto más ambicioso, un portal de portales: el *Servidor Web de Literatura Española «Parnaseo»* (Canet 2014: 13). Además de *LEMIR*, que albergaba entonces, como hemos visto, las secciones de *Memorabilia* y *Tirant*, se alojó en *Parnaseo* un nuevo portal: *Ars Theatrica*, dirigido por el propio Canet junto a Evangelina Rodríguez y Josep Lluís Sirera, y centrado en el teatro español. En paralelo a estas dos grandes secciones, *Parnaseo* se planteó mantener charlas en tiempo real a través de la propuesta de *Parnachat*,²⁶ así como ediciones facsimilares, enlaces de interés e, inicialmente, dos bases de datos de producción propia con una estrecha relación: *Imprenta valenciana siglo XVI* y *Tipobibliografía valenciana siglos XV y XVI*.²⁷ La primera de ellas aloja un repertorio abreviado bibliográfico de la producción impresa en Valencia durante el siglo XVI,²⁸ mientras que la

24.- El portal *LEMIR* fue una evolución de un *gopher* —un servicio de Internet textual que funciona mediante menús jerárquicos— que había montado previamente y que alojaba textos medievales españoles (Canet 2014: 12). Se puede consultar una de las primeras versiones del portal en <<https://web.archive.org/web/19970607204833/http://www.uv.es/~lemir/>> [consulta: 06/04/2023].

25.- Este *Boletín de Literatura Sapiencial* recibe el nombre de *Exemplaria* en su número 0, pero ya en el 1 pasa a llamarse *Memorabilia*, tal y como se le conoce hasta hoy.

26.- A mediados de los 90, el chat era un sistema de comunicación textual en tiempo real muy utilizado. Se basaba en un protocolo de comunicación, el IRC, que permitía establecer comunicaciones privadas y grupos de discusión —salas—. Con el tiempo, surgieron otras alternativas, como el Microsoft Messenger, que también han sido relegadas por las actuales: Whatsapp y Telegram.

27.- Se puede consultar la versión de *Parnaseo* del 5 de diciembre de 1998 en <<https://web.archive.org/web/19981205234850/https://parnaseo.uv.es/>>.

28.- La base de datos *Imprenta valenciana siglo XVI*, entre otras cosas, ha permitido documentar el proceso de retroceso del valenciano respecto al castellano en la producción literaria de la época y que los textos valencianos publicados eran mayormente de poesía religiosa y profana (Canet 2004: 20-21).

segunda²⁹ lo desarrolla para ofrecer la catalogación tipográfica completa de la imprenta valenciana de los siglos XV y XVI.³⁰

En un momento en el que los facsímiles digitales no existían y las bibliotecas estaban en pleno proceso de informatización de su catálogo completo, la recolección de los datos para la confección de la *Tipobibliografía valenciana siglos XV y XVI* fue un trabajo complejo (Hernández Royo 1994: 43-44). Desde entonces, esta base de datos ha estado en continuo crecimiento y actualmente aloja más de 900 fichas. Dispone de una intuitiva interfaz organizada en tres niveles de detalle: el buscador, el resultado del filtrado y el contenido propio de cada ficha. Esta estudiada disposición de elementos en una única página agiliza el acceso a cada ficha individual cuando se tienen que consultar varias de ellas, al no tener que volver a una página anterior y recargarla para obtener de nuevo el listado. Además, el contenido de cada una de las fichas es hipermedial, está enriquecido con imágenes que acompañan las descripciones textuales: «Cada ficha permite la consulta y ampliación de las imágenes de las portadas, preliminares, colofones, grabados, letras capitales, escudos, filigranas y toda la información editorial pertinente» (Canet 2019: 456).

Desde su creación, *Parnaseo* no ha dejado de crecer con la creación de nuevas bases de datos, como la de *Carteles teatrales valencianos del siglo XIX*, dirigida por Canet en 2007,³¹ así como con la incorporación de otras revistas electrónicas, como la histórica *Celestinesca*,³² o la creación de nuevos portales, entre los que destacan los más recientes coordinados por Marta Haro Cortés: *Aula Medieval* y *Portal Celestinesco*.

Beltrán, Canet y Haro, en un momento en el que la publicación en papel era la norma, impulsaron la creación de revistas digitales y bases de datos que, más de veinte años después, han seguido publicándose y renovándose. Fueron unos visionarios que supieron aprovechar las posibilidades que brindaba la red como medio de difusión. El tiempo les ha dado la razón. *Parnaseo* ha continuado creciendo y se ha convertido en un servidor web de referencia en el mundo filológico hispánico en el que, actualmente, se alojan diversos portales temáticos, bases de datos, publicaciones electrónicas y su propia línea editorial.

Por su parte, SAGRARIO LÓPEZ POZA vio en las humanidades digitales la oportunidad para acceder con mayor facilidad a datos sobre los que fundamentar sus investigaciones, así como un mecanismo para difundir los resultados que generaba.³³ Por ello, creó el *Seminario interdisciplinar para el estudio de la Literatura Áurea Española* (SIELAE):

29.– Aunque se puso en línea posteriormente, se empezó a desarrollar a principios de los 90. El proyecto figura activo desde 2010 bajo la dirección de Marta Haro y con José Luis Canet como responsable de la sección (Canet 2019: 455).

30.– «Los materiales utilizados en esta base de datos corresponden a tres tesis doctorales realizadas por Purificación Hernández Royo, *La imprenta valenciana de la familia Mey-Huete en el siglo XVI: Producción y Tipografía*, Universitat de València, Facultat de Filologia, octubre de 1994; Gloria Irún de Sojo, *Catálogo gráfico-descriptivo de la Imprenta del Molino de la Rovella: s. XVI*, Universitat de València, Facultat de Filologia, enero de 1995 y Diego A. Romero Lucas, *Catálogo gráfico descriptivo de la Imprenta en Valencia (1473-1530)*. Universitat de València, Facultat de Filologia, 16 de febrero de 2005; todas ellas dirigidas por el prof. José Luis Canet. También se incluyen otros materiales recogidos posteriormente». (*Tipobibliografía* 2023).

31.– Para otras bases de datos, véase <<https://parnaseo.uv.es/Bases.htm>>.

32.– Bajo la dirección de Joseph T. Snow, que se venía editando en papel desde 1977 y en 2003 pasó a formar parte también de *Parnaseo*.

33.– «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

La idea de crear el SIELAE (Seminario interdisciplinar para el estudio de la Literatura Áurea Española) 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 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 (López Poza & Saavedra Places 2014: 285-286).

La coexistencia de filólogos e informáticos en un mismo equipo de trabajo permitió abordar metodologías para el estudio filológico que habrían sido inconcebibles de otra manera, avaladas por la propia circunstancia de que las convocatorias estatales de proyectos incentivaban que se estableciesen ese tipo de colaboraciones. La primera línea de investigación que demostró los beneficios de un grupo multidisciplinar de tales características fue la relativa a las *Relaciones de sucesos*:

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 (Pena Sueiro & Álvarez García 2014: 338).

López Poza confió a Nieves Rodríguez Brisaboa³⁴ la parte tecnológica de sus investigaciones. Ambas formaron un equipo en 1994 (Rodríguez Brisaboa *et al.*: 33) que perduró en el tiempo y que, en última instancia, fue el responsable de la aparición de un portal de referencia en el mundo filológico tan relevante como la *Biblioteca Digital Siglo de Oro* (BIDISO).

La primera incursión en la red de López Poza comenzó en septiembre de 1996 con la publicación de la web de *Literatura Emblemática Hispánica*.³⁵ Su objetivo era alojar la producción del equipo internacional que había reunido y que estaba compuesto por investigadores de las ramas de la computación y de la filología.³⁶ Esta primera versión ofrecía información sobre sus líneas de trabajo y proyectos, bibliografía específica, noticias de interés y enlaces a otras páginas académicas de interés. Justo un mes después, junto con Nieves Pena Sueiro, ponía en línea otra página: el boletín de *Relaciones de sucesos españolas*

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.» (López Poza & Saavedra Places 2014: 286).

34.- Esta colaboración permitió a Rodríguez Brisaboa la creación del Laboratorio de Bases de Datos, «cuya línea de investigación aplicada fundamental fue, como no, la de las Bibliotecas Digitales» (Rodríguez Brisaboa *et al.* 2019: 35).

35.- Se puede consultar la versión de 1998 en la que se lee, a pie de página, que está en línea desde septiembre de 1996, en <<https://web.archive.org/web/19981202185628/http://rosalia.dc.fi.udc.es/emblematica/>> [consulta: 04/04/2023].

36.- Se pueden consultar los miembros del grupo Emblemática en la siguiente dirección: <<https://web.archive.org/web/19991022031636/http://rosalia.dc.fi.udc.es/emblematica/Presentacion.html>> [consulta: 04/04/2023].

en la *edad moderna*,³⁷ cuyo primer número recopilatorio se publicaba en enero de 1997.³⁸ «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» (Pena Sueiro & Álvarez García 2014: 339). En paralelo a la creación de estas páginas, el equipo estaba embarcado en recopilar y digitalizar ejemplares de relaciones de sucesos dispersos a lo largo de la geografía española para almacenarlos en una base de datos, un trabajo que suponía todo un reto tanto por la cantidad de material que debían manejar como por el medio físico en el que se encontraba.³⁹ Todo el contenido recopilado se pudo consultar públicamente en Internet a partir de 2001 (Pena Sueiro 2017: 76-79).

López Poza, con la creación del SIELAE, apostó por los equipos multidisciplinares como una nueva forma de abordar las investigaciones humanísticas que, hasta ese momento, habían estado confinadas únicamente al ámbito filológico. Ese nuevo enfoque les permitió avanzar en paralelo en sus investigaciones e intercambiar conocimiento que ha beneficiado tanto a las ciencias humanas como a las ciencias computacionales. Marcó un camino a seguir que, aunque este modelo de interacción y/o cooperación aún no se ha consolidado, es cada vez es más habitual.

El uso de Internet como plataforma de difusión de conocimiento también permitió ampliar el concepto clásico de fondo bibliográfico con la aparición de las *bibliotecas digitales* «a texto completo, o mejor dicho a imagen completa de los libros, diarios, mapas, grabados, etc., capaz de ser visualizados mediante cualquier navegador» (Canet 2000: 71), un concepto que implica una evolución respecto a la consulta del catálogo vía web. En este caso, además de la ficha bibliográfica, se proporciona acceso a una edición digital de la obra, siguiendo el modelo iniciado por el *Proyecto Gutenberg*.⁴⁰ Algunas de las primeras bibliotecas españolas que se aventuraron a ofrecer este tipo de acceso fueron la *Biblioteca Digital de Catalunya* (BDC), la *Biblioteca Digital de la Universitat de València* (BDUV) y la *Biblioteca Virtual Miguel de Cervantes* (BVMC).

La BDC fue impulsada por el *Concorci de Biblioteques Universitàries de Catalunya* (CBUC)⁴¹ con la finalidad de «proporcionar un conjunto nuclear de información electrónica interdisciplinaria para la totalidad de la comunidad universitaria e investigadora de Catalunya independientemente de donde estas personas ejerzan su actividad» (Anglada & Comellas 2000: 243). El proyecto comenzó a gestarse en 1997 y en 1999 se publicaba su primera versión.⁴² A diferencia del Proyecto Gutenberg, que tenía como fin ofrecer

37.– El boletín, más tarde, pasó a llamarse BORESUS. Aunque la primera versión capturada por Internet Archive es de febrero de 1999, en la portada se indica que está en línea desde octubre de 1996. Se puede consultar en <<https://web.archive.org/web/19990219220356/http://rosalia.dc.fi.udc.es/BORESU/>> [consulta: 04/04/2023].

38.– La sección *Relaciones de sucesos* de BIDISO alberga el repositorio de los boletines publicados desde el número 1.

39.– Se comenzó el proceso con la digitalización de microfilms (Pena Sueiro 2017: 76).

40.– «En juillet 1971, Michael Hart crée le Projet Gutenberg pour diffuser gratuitement sous forme électronique les oeuvres littéraires du domaine public. Un projet longtemps considéré par ses détracteurs comme impossible à grande échelle. Site pionnier à tous égards, le Projet Gutenberg est à la fois le premier site d'information sur un réseau encore embryonnaire et la première bibliothèque numérique. Michael numérise lui-même les cent premiers livres» (Lebert 2005: 3).

41.– Está formado por las bibliotecas de las universidades públicas catalanas y la Biblioteca de Catalunya.

42.– La primera versión de la BDC daba acceso a «BDA Aranzadi (sobre derecho español), Business Source Elite, Econlit, ERIC, Inside, Mathscinet, Medline y The Serials Directory. También se cuenta con la suscripción a las revistas de

versiones digitales de material impreso, su objetivo inicial era proporcionar únicamente acceso a diversas bases de datos y revistas electrónicas con la intención de abaratar costes con la suscripción conjunta, aunque «durant el període que s'inicia el 2005 la BDC ha entrat de ple en l'àmbit de les compres/subscripcions a llibres electrònics» (Anglada & Comellas 2010: 10). Actualmente, aloja más de 53 000 recursos electrónicos, con la posibilidad de buscar en su amplio catálogo tanto desde su web como desde los buscadores de las instituciones participantes.

En ese momento en el que nacían las bibliotecas digitales, Canet ocupaba el cargo de *Director del Servei d'Informació Bibliogràfica de la Universitat de València*. Como defensor de la difusión de la información en Internet, en 1999 y con motivo de la celebración del V Centenario de la universidad, propuso la creación del proyecto *Biblioteca Digital*. Su objetivo era poner en la red materiales digitalizados para fines de investigación y «mostrar los fondos valiosos depositados en su Biblioteca Histórica, enlazando así con otro de los proyectos aprobados, *Thesaurus*, encargado de la catalogación y difusión del patrimonio de la Universitat» (Canet 2000: 72).⁴³ Al año siguiente, se ponía en marcha

el proyecto de digitalización del fondo antiguo a texto completo denominado *Somni* (sueño). Este término deriva del título del incunable *El somni de Johan Johan* de Jaime Gazull, ejemplar único en el mundo, impreso en Valencia, en 1497 y que se conserva en la *Biblioteca Històrica de la Universitat de València* (Millás Mascarós & Escriche Soriano 2017: 4).

La colección *Somni* se integró en el 2008 en el *Repositori d'Objectes Digitals per a l'Ensenyament, la Recerca i la Cultura* (RODERIC), de tipo *Open Access*, y en el 2010 formó parte de *Europeana Regia*.⁴⁴ Dado que las primeras digitalizaciones se habían llevado a cabo a partir de microfilms, con motivo de la participación en este proyecto, se realizó de nuevo una captura, pero esta vez directamente de las obras, con el fin de disponer de imágenes de mayor resolución (Millás Mascarós & Escriche Soriano 2017: 5). Actualmente, la colección cuenta con más de 8 000 reproducciones digitales, entre las que se encuentran colecciones de carteles, libretos de fallas, manuscritos, mapas, publicaciones periódicas y libros desde el siglo XV al XX.

Por su parte, la *Biblioteca Virtual Miguel de Cervantes* (BVMC) aparecía en Internet en 1999 como un espacio puramente virtual, es decir, no estaba unido a una biblioteca física. Era un proyecto impulsado por la Universitat d'Alacant y patrocinado por el Banco de Santander y la Fundación Botín (Rovira Soler & Rovira-Collado 2019: 54). Al igual que el resto de bibliotecas digitales, abrió la posibilidad de acceder a las obras de forma totalmente gratuita a través de Internet, hecho que llevó a tratar con especial atención los derechos de autor (Rovira Soler 2001: 68). Desde sus inicios, la BVMC ha sido un portal

Academic Press (IDEAL). Además, se da acceso a diversas bases de datos gratuitas catalanas, producidas por universidades y centros de investigación» (Anglada & Comellas 2000: 242).

43.- Los materiales que se querían almacenar inicialmente eran prensa periódica antigua (s. XVIII y XIX), libros del s. XV, XVI, XVII, XVIII y XIX, manuscritos, carteles/grabados/mapas y la colección de monedas de la Universitat (Canet 2000: 73).

44.- *Europeana Regia* fue un proyecto europeo para digitalizar el fondo bibliográfico de manuscritos de la Edad Media y el Renacimiento, en concreto, la Biblioteca Carolingia, la Biblioteca de Carlos V y la Biblioteca de los Reyes de Aragón y Nápoles.

que ha aportado ideas innovadoras a las humanidades digitales, como la transcripción de las fuentes digitalizadas o la creación de versiones sonoras de las obras. Además, dispone de una sección específica para apoyar la creación de nuevas formas de uso de los datos digitales y del que han salido herramientas de suma utilidad como, por ejemplo, el analizador de versos, que, a partir un poema, nos indica de forma automática su cadencia, las sinalefas, los acentos y el número de sílabas. Todos sus datos son accesibles en abierto y se ha ido adaptando a los estándares según ha ido pasando el tiempo. Actualmente, tiene la base de datos en el estándar *RDF* y el vocabulario *RDA*⁴⁵ (Candela *et al.* 2018), que se pueden consultar a través de un punto de acceso *SPARQL*.⁴⁶ La *BVMC*, en cualquier caso, no solo se limita a proporcionar acceso abierto a sus datos, sino que también enriquece su contenido a partir de repositorios externos, como es el caso de las fichas de autor, que obtienen parte de su información de Wikidata.⁴⁷ Es toda una demostración de las posibilidades de la web semántica⁴⁸ y del trabajo colaborativo para la creación de conocimiento.

En definitiva, los proyectos filológicos que, en el último cuarto del siglo pasado, utilizaron la informática y se adentraron en la red de redes sentaron las bases sobre las que se han erigido las investigaciones actuales, cuya presencia en el universo digital ya resulta indispensable. Ciertos humanistas e informáticos comenzaron el camino de las humanidades digitales en España, abrazando la tecnología y la colaboración interdisciplinar como la norma y no la excepción. Impulsaron la digitalización y la difusión pública de contenidos. Su visión de futuro ha permitido que, en pleno siglo XXI, tengamos revistas científicas digitales en abierto, repositorios que se puedan consultar sin suscripción y fac-símiles digitales accesibles a golpe de clic. Las humanidades son digitales gracias a ellos.

45.– *RDF* y *RDA* son estándares que facilitan el intercambio de información entre ordenadores definiendo el formato de los datos.

46.– *SPARQL* es un lenguaje informático para realizar consultas sobre datos definidos en *RDF*.

47.– *Wikidata* es una base de datos que permite la edición en abierto de forma colaborativa <<https://www.wikidata.org/>>. Está pensada para utilizarla desde otras aplicaciones, a diferencia de la *Wikipedia*, que presenta la información para lectores humanos.

48.– Pascal Hitzler ha resumido en un artículo la historia y el uso de la web semántica (2021).

Bibliografía citada

- [Admyte] Francisco Marcos Marín, Charles Faulhaber, & Ángel Gómez Moreno, *Biblioteca electrónica Admyte. Archivo Digital de Manuscritos y Textos Españoles*, Hispanic Seminary of Medieval Studies, University of Wisconsin. <<https://www.admyte.com>> [consulta: 29/12/2022].
- [AEPIA] *Asociación Española para la Inteligencia Artificial*. <<https://www.aepia.org>> [consulta: 02/04/2023].
- ALVAR EZQUERRA, Manuel, «Obtención automática de índices de rimas y de sufijos», *Revista de dialectología y tradiciones populares*, 32 (1976), pp. 35-42.
- ANGLADA, Lluís, & Núria Comellas, «La Biblioteca Digital de Catalunya: oportunidades, opciones y estrategias en la adquisición compartida de información electrónica», en *Primeras Jornadas de bibliotecas digitales*, ed. Nieves R. Brisaboa et al., Valladolid, Universidad de Valladolid, 2000, pp. 237-248.
- , «Biblioteca Digital de Catalunya: 10 anys d'activitats», en *12 jornades catalanes d'informació i documentació*, Barcelona, 2010.
- ARIZA, Manuel, et al., «Atlas lingüísticos plurilingües con ordenadores electrónicos», *Boletín del Centro de Cálculo de la Universidad de Madrid*, 23 (1973), pp. 12-15.
- ARROYO GALÁN, Luis, *100 años de Informática y Telecomunicaciones. España siglo XX*, Madrid, Fundación Rogelio Segovia para el Desarrollo de las Telecomunicaciones, 2005.
- BADIA, Lola, José M. BLECUA, Glòria CLAVERIA, Josep PUJOL, Amadeu SOBERANAS & Joan TORRUELLA (1995), *Els cançoners catalans medievals. Concordances*, Bellaterra, Fundació «La Caixa» - Seminari de Filologia i Informàtica, Universitat Autònoma de Barcelona, 10 vols. [microfichas].
- [BETA] Charles B. Faulhaber, et al., *Biblioteca Española de Textos Antiguos*, en *Philobiblon*, Berkeley, The Bancroft Library - University of California Berkeley. <<http://sunsite.berkeley.edu/Philobiblon/phhmbe.html>> [consulta: 29/12/2022].
- [BIDISO] Sagrario López Poza, & Nieves Pena Sueiro, *Biblioteca Digital Siglo de Oro*, Universidad da Coruña. <<https://bidiso.es>> [consulta: 01/04/2023].
- [BIPA] Ralph DiFranco, et al., *Bibliografía de Poesía Áurea*, en *PhiloBiblon*, <https://bancroft.berkeley.edu/philobiblon/bipa_es.html> [consulta 03/02/2023]
- [BITAGAP] Arthur L-F. Askins, et al., *Bibliografía de Textos Antigos Galegos e Portugueses*, en *Philobiblon*, Berkeley, The Bancroft Library - University of California. <https://bancroft.berkeley.edu/philobiblon/bitagap_es.html> [consulta 03/02/2023].
- [BITECA] Vicenç Beltran, et al., *Bibliografía de Textos Antics Catalans*, en *Philobiblon*, Berkeley, The Bancroft Library - University of California Berkeley. <<http://sunsite.berkeley.edu/Philobiblon/phhmbi.html>> [consulta: 06/07/2022].
- [BOOCT] Beatrice Jorgensen Concheff, *Bibliography of Old Catalan Texts*, Madison, Hispanic Seminar of Medieval Studies, 1985.
- [BOOST] Charles B. Faulhaber, *Bibliography of Old Spanish Texts* [1975], Madison, Hispanic Seminary of Medieval Studies, 1984.
- BOOTH, Andrew Donald, & William N. LOCKE, «Historical Introduction», en *Machine Translation of languages: fourteen essays*, ed. A. D. Booth & W. N. Locke, 1955, pp. 1-14.
- BUSA, Roberto, «The Annals of Humanities Computing: the *Index Thomisticus*», *Computer and Humanities*, 14/2 (1980), pp. 83-90. <<https://doi.org/10.1007/BF02403798>>.
- [BVMC] *Biblioteca Virtual Miguel de Cervantes*, Universidad de Alicante, Generalitat Valenciana y Biblioteca Valenciana. <<https://www.cervantesvirtual.com/>> [consulta: 02/04/2023].
- CABRÉ CASTELLVÍ, M. Teresa, «Bernard Quemada, una figura cabdal en la lingüística francesa (1926-2018)», *Estudis Romànics*, 41 (2019), pp. 691-725.

- CAMPO, Ignacio del, *et al.*, «El análisis sintáctico automático como una ayuda para la elaboración del diccionario», *Boletín del Centro de Cálculo de la Universidad de Madrid*, 23 (1973), pp. 16-31.
- CANDELA, Gustavo, *et al.*, «Migration of a Library Catalogue into RDA Linked Open Data», *Semantic Web Journal*, 9/4 (2018), pp. 481-491. <<http://dx.doi.org/10.3233/SW-170274>>.
- CANET, José Luis, «La nueva Biblioteca Digital de la Universitat de València», en *Primeras Jornadas de bibliotecas digitales*, ed. Nieves R. Brisaboa *et al.*, Valladolid, Universidad de Valladolid, 2000, pp. 71-78.
- , «Literatura i impremta durant el segle XVI a València», en *Escriptors valencians de l'Edat Moderna*, Valencia, Acadèmia Valenciana de la Llengua, 2004, pp. 19-32.
- , «Reflexiones sobre las humanidades digitales», *Humanidades Digitales: desafíos, logros y perspectivas de futuro. Janus*, Anexo 1 (2014), pp. 11-20.
- , «Tipobibliografía valenciana de los siglos XV y XVI», *Historias Fingidas*, 7 (2019), pp. 455-458. <<http://dx.doi.org/10.13136/2284-2667/139>>.
- CASTAÑOS ALÉS, Enrique, *Los orígenes del arte cibernético en España: el seminario de Generación Automática de Formas Plásticas del Centro de Cálculo de la Universidad de Madrid: (1968-1973)*, Alicante, Biblioteca Virtual Miguel de Cervantes, 2000.
- [CORDE] *Corpus Diacrónico del Español*, Real Academia Española. <<https://corpus.rae.es/cordecnet.html>> [consulta: 15/04/2023].
- [CORPES] *Corpus del Español del Siglo XXI*, Real Academia Española. <<https://apps2.rae.es/CORPES>> [consulta: 15/04/2023].
- [CREA] *Corpus de Referencia del Español Actual*, Real Academia Española. <<https://corpus.rae.es/creanet.html>> [consulta: 15/04/2023].
- FAULHABER, Charles B., «PhiloBiblon: pasado y futuro», *Incipit*, XXIX (2009), pp. 191-200.
- , «PhiloBiblon, Information Technology, and Medieval Spanish Literature: A Balance Sheet», en *Humanistats a la xarxa: món medieval*, ed. Lourdes Soriano *et al.*, 2014, pp. 15-43.
- FRANCIS, W. Nelson, «A Standard Corpus of Edited Present-Day American English», *College English*, 26/4 (1965), pp. 267-273.
- FRANCIS, W. Nelson, & Henry KUCERA, *Brown corpus manual*, Brown University, 1979.
- GARCÍA CAMARERO, Ernesto, «Noticias», *Boletín del Centro de Cálculo de la Universidad Complutense*, 28 (1976), pp. 102.
- GARIJO, Francisco Javier, & María Felisa VERDEJO, «Preprocesador de Algol 60 por el método A.E.D.», *Boletín del Centro de Cálculo de la Universidad de Madrid*, 22 (1973), pp. 48-62.
- HERNÁNDEZ ROYO, Pura, *La imprenta valenciana de la familia Mey-Huete en el siglo XVI: producción y tipografía*, Valencia, Universitat de València, 1994 [tesis doctoral].
- HITZLER, Pascal, «A Review of the Semantic Web Field», *Communications of the ACM*, 64/2 (2021), pp. 78-83. <<http://dx.doi.org/10.1145/3397512>>.
- [HSMS] *Hispanic Seminary of Medieval Studies*, University of Wisconsin-Madison. <<http://www.hispanicseminary.org/>> [consulta: 12/04/2023].
- LEBERT, Marie, *Le Project Gutenberg (1971-2005)*, Troisième symposium international sur les études françaises valorisées par les technologies: langages et dialogues interculturels, Toronto, Canada, 2005.
- LÓPEZ, Aramis, & Jaime MUNARRIZ, *El Centro de Cálculo de la Universidad de Madrid (1968-1973): ciencia, arte y creación computacional*, Madrid, Ediciones Complutense, 2021.
- LÓPEZ POZA, Sagrario, «Humanistas y Humanidades digitales. Trayectoria y proyección en la Filología española», en *Humanidades y Humanismo. Homenaje a María Pilar Cuartero*, ed. Aurora Egido *et al.*, Zaragoza, Institución Fernando el Católico, 2020, pp. 15-43.

- LUCÍA MEGÍAS, José Manuel, «Los nuevos filólogos del siglo XXI: la literatura medieval hispánica en la red», en *Actas XIII Congreso AHML*, ed. José Manuel Fradejas *et al.*, Valladolid, Ayuntamiento de Valladolid y Universidad de Valladolid, 2010, pp. 1233-1254.
- MARCOS MARÍN, Francisco A., «Posibilidad y dificultades de la traducción automática», *Filología Moderna*, 42 (1971), pp. 313-327.
- , *Informática y humanidades*, Madrid, Gredos, 1994.
- , «Historia humana de la lengua española y su computación», *Studies in Hispanic and Lusophone Linguistics*, 2/2 (2009), pp. 387-416. <<https://doi.org/10.1515/SHLL-2009-1057>>.
- MILLÁS MASCARÓS, Elisa, & Margarita ESCRICHE SORIANO, «La Biblioteca Digital de la Universitat de València. difusión y preservación de fondos históricos», *RUIDERAe: Revista de Unidades de Información*, 12 (2017).
- MENABREA, Luigi Federico, «Sketch of the Analytical Engine invented by Charles Babbage», en *Scientific Memoirs*, 3, London, Richard Taylor, 1843, pp. 666-731 [1ª ed. en francés: «Notions sur la machine analytique de M. Charles Babbage», *Bibliothèque Universelle de Genève*, 41 (1842), pp. 352-376].
- MEYA, Montserrat, «Editorial», *Procesamiento del Lenguaje Natural*, 1 (1983), pp. 1-4.
- NITTI, John J., «Computers and the Old Spanish Dictionary», *Computers and the Humanities*, 12-1/2 (1975), pp. 43-52.
- [Parnaseo] José Luis Canet, & Marta Haro, *Servidor Web de Literatura Española*, Universidad de Valencia. <<https://parnaseo.uv.es>> [consulta: 01/04/2023].
- PÊCHEUX, Michel, *Hacia el análisis automático del discurso*, Madrid, Gredos, 1978.
- PENA SUEIRO, Nieves, & Sandra ÁLVAREZ GARCÍA, «El Catálogo y Biblioteca digital de relaciones de sucesos: bases de datos bibliográficas, textos e imágenes», *Humanidades Digitales: desafíos, logros y perspectivas de futuro. Janus*, Anexo 1 (2014), pp. 285-303.
- PENA SUEIRO, Nieves, «El portal BIDISO: pasado, presente y futuro inmediato. Un ejemplo de evolución en aplicaciones de las HD», *Studia Aurea*, 11 (2017), pp. 73-93. <<https://doi.org/10.5565/rev/studiaaurea.264>>.
- RALSTON, Anthony, & Edwin D. REILLY, *Encyclopedia of Computer Science and Engineering*, 2ª ed., New York, Van Nostrand Reinhold, 1983.
- [RODERIC] *Repositori d'Objectes Digitals per a l'Ensenyament, la Recerca i la Cultura*, Universitat de València, <<https://roderic.uv.es/>> [consulta: 15/04/2023].
- RODRÍGUEZ BRISABOA, Nieves, *et al.*, «Sagrario López Poza, humanista digital», en *Festina Lente. Augusta empresa correr a espacio. Studia in honorem Sagrario López Poza*, ed. Carlota Fernández Travieso & Nieves Pena Sueiro, A Coruña, Servizo de Publicacións Universidade da Coruña, 2019, pp. 33-42.
- ROJO, Guillermo, «Corpus textuales del español», en *Enciclopedia de Lingüística Hispánica*, ed. Javier Gutiérrez, Londres, Routledge, 2015, pp. 285-296.
- ROJO, Guillermo, «Citius, maius, melius: del CREA al CORPES XXI», en *Lingüística de corpus y lingüística histórica iberorrománica*, ed. Johannes Kabatek & Carlota de Benito Moreno, Berlin, De Gruyter, 2016, pp. 197-212.
- ROVIRA SOLER, José Carlos, «Sobre los textos y sus derechos en la Biblioteca Virtual Miguel de Cervantes en la Universidad de Alicante», *Métodos de Información*, 8/44 (2001), pp. 67-70. <<https://doi.org/10.5557/%25x>>.
- ROVIRA SOLER, José Carlos, & José ROVIRA-COLLADO, «La literatura en español en Internet: veinte años de la creación de la Biblioteca Virtual Miguel de Cervantes», *Mi Biblioteca*, 58 (2019), pp. 54-59.

- SÁNCHEZ SÁNCHEZ, Mercedes, & Carlos DOMÍNGUEZ CINTAS, «El banco de datos de la RAE. CREA y CORDE», *Per Abbat: boletín filológico de actualización académica y didáctica*, 2 (2007), pp. 67-70.
- [SEPLN] *Sociedad Española para el Procesamiento de Lenguaje Natural*. <<http://www.sepln.org>> [consulta: 01/04/2023].
- TORRUELLA, Joan, *La rima en la lírica medieval (estudi mètric del cançoner L)*, Bellaterra, Universitat de Autònoma de Barcelona, 1992.
- TORRUELLA, Joan, «Arxiu informatitzat de textos catalans medievals», en *Actes del novè Col·loqui Internacional de Llengua i Literatura Catalanes vol. 2*, Barcelona, Abadia de Montserrat, 2003, pp. 239-252.
- TORRUELLA, Joan, & Jeremy N. H. LAWRENCE, «Un projecte d'arxiu informatitzat de textos catalans medievals: algunes normes», *Llengua i Literatura*, 3 (1988-1989), pp. 481-506.
- TURING, Alan, «On Computable Numbers, with an Application to the Entscheidungsproblem», en *Proceedings of the London Mathematical Society*, s2/42 (1937), pp. 230-265. <<https://doi.org/10.1112/plms/s2-42.1.230>>.
- VANHOUTTE, Edward, «The Gates of Hell: History and Definition of Digital | Humanities | Computing», en *Defining Digital Humanities: A Reader*, ed. Melissa Terras *et al.*, New York Routledge, 2013, pp. 119-156.
- VERDEJO, María Felisa, «Un robot capaz de dialogar en castellano», *Boletín del Centro de Cálculo de la Universidad Complutense*, 29 (1976), pp. 1-16.

