



PROGETTO
MAMBRINO

HISTORIAS FINGIDAS



Humanidades Digitales y literatura medieval española: la integración de Transkribus en la base de datos COMEDIC

Nuria Aranda García

(École Normale Supérieure de Lyon)*

Abstract

La base de datos COMEDIC (Catálogo de obras medievales impresas en castellano) nace en 2012 con el objetivo de analizar la recepción y transmisión de los textos literarios medievales en la imprenta quinientista. Los avances en la aplicación de herramientas de las Humanidades Digitales y las sucesivas renovaciones del proyecto lo han conducido al empleo de estos útiles en la investigación que desarrolla. En el presente artículo se hace un estado de la cuestión y una proyección sobre los objetivos que se pretenden conseguir mediante la aplicación del programa Transkribus.

Palabras clave: COMEDIC; Humanidades Digitales; Transkribus; ediciones digitales académicas; literatura medieval española

The database COMEDIC (Catálogo de obras medievales impresas en castellano) was created in 2012 with the aim of analysing the reception and transmission of medieval literary texts in the 16th century print. Advances in the application of Digital Humanities tools and the successive renewals of the project have led it to apply these tools to its research data. This paper offers a state of the art and a projection of the goals to be achieved in the application of the Transkribus software.

Keywords: COMEDIC; Digital Humanities; Transkribus; Digital Scholar Editions; Spanish Medieval Literature



* Este trabajo se ha realizado en el marco del Proyecto de Investigación PID2019-104989GB-I00, financiado por MCIN/AEI/10.13039/501100011033, y se inscribe en el grupo investigador CLARISEL, que cuenta con la participación económica del Departamento de Ciencia, Universidad y Sociedad del Conocimiento del Gobierno de Aragón.

Introducción

Las Humanidades Digitales debieron su nacimiento al deseo por potenciar la divulgación y visibilidad de los resultados científicos de esta disciplina mediante el apoyo proporcionado por el medio digital. Este no solo les ha servido para ganar reconocimiento frente a otras disciplinas en el panorama investigador actual, sino que las ha empujado a la interdisciplinariedad y a la formación de equipos de trabajo donde se dan la mano especialistas de estudios diversos y donde es necesaria la comunicación con grupos de informáticos (Spence, 2014). Pese a que estas reflexiones, sencillas en su conceptualización teórica, han comenzado a aplicarse en la práctica en muchos trabajos, las Humanidades Digitales apenas han comenzado a despertar en nuestro país, aunque el balance que puede realizarse hasta el momento es bastante positivo e indica que caminamos hacia la buena dirección (González Blanco-García, 2013; López Poza, 2015)

En el ámbito hispánico, lejano queda aparentemente en el tiempo BOOST (*The Bibliography of Old Spanish Texts*), actualmente la base de datos *PhiloBiblon*, considerado como primer proyecto de Humanidades Digitales orientado al tratamiento de textos en español, y desarrollado por Charles Faulhaber y Francisco Marcos Marín en la década de los 70 en el seno del *Hispanic Seminar of Medieval Studies* de la Universidad de Wisconsin¹. Las Humanidades comenzaron a incorporar herramientas informáticas en España en la década de los 80, momento en que puede situarse la labor de los pioneros de nuestro país, cuyos trabajos sirvieron como punto de partida para el desarrollo de los primeros proyectos especialmente en los departamentos de filología (Canet, 2014). El incremento de subvenciones por parte de organismos públicos e instituciones privadas potenció las Humanidades Digitales en la década de los 90, y esto se tradujo en una proliferación de trabajos cuya subsistencia y sostenibilidad se debió y se debe a esta financiación y a la proporcionada a los proyectos dentro del

¹ Son muchos los trabajos que actualmente ofrecen un estado de la cuestión sobre las Humanidades Digitales en España, sirvan como ejemplo Rojas Castro (2013), Spence y González-Blanco García (2014), López Poza y Pena Sueiro (2014), Morrás y Rojas Castro (2015), López Poza (2015; 2019) Toscano, Rabadán, Ros y González-Blanco García (2020) y Hernández Lorenzo (2020).

programa estatal de I+D+I (Toscano, Rabadán, Ros y González-Blanco García, 2020). Sirvan como ejemplos representativos de esta etapa inicial ADMYTE de Charles Faulhaber y Francisco Marcos Marín (1992) (Faulhaber y Marcos Marín, 1992) y el portal LEMIR de José Luis Canet (1995) (Canet, 2014, 12). A todo lo anterior se sumaron los fondos FEDER, que promovieron una colaboración interdisciplinar entre investigadores de distintas áreas humanísticas y la incorporación de grupos de informáticos (López Poza, 201, 138) que culminó en la creación de bases de datos de autores y obras, la edición y estudio de determinados géneros y la creación de bibliotecas digitales de corpus específicos que en tiempos más recientes han incorporado el lenguaje de marcado (López Poza, 2015: 3).

Con la llegada del nuevo milenio, y fundamentalmente a partir de 2005 y 2006, comienzan ya a proliferar las publicaciones con teorizaciones sobre esta nueva disciplina y sobre la edición de textos electrónicos y el desarrollo de ediciones digitales (Hernández Lorenzo, 2020, 566-568), mientras la etiqueta «Humanidades Digitales» ya queda perfectamente acuñada en España. Sustituta de la lejana «Informática humanística», que respondía a la inseguridad de estos investigadores hacia los aspectos técnicos, el nuevo sintagma pone en primer lugar a la figura del humanista que debe adquirir las destrezas de las nuevas tecnologías (López Poza, 2019, 127). Desde entonces, las Humanidades Digitales han seguido un proceso de consolidación hasta institucionalizarse en 2011, momento en que el panorama inicial de falta de comunicación entre proyectos afines e iniciativas inconexas se transforma gracias a la sucesión de eventos y encuentros relacionados con las Humanidades Digitales que culminan en la creación de la asociación *Humanidades Digitales Hispánicas. Sociedad Internacional* (HDH)². Ahora las publicaciones van a centrarse en aplicaciones concretas del medio digital en los estudios humanísticos, va a potenciarse la formación en Humanidades Digitales mediante su inclusión en los planes de estudios y van a sucederse los encuentros con especialistas de este campo (Hernández Lorenzo, 2020, 572-579). Proliferan las bases de datos de autores, obras y géneros en el ámbito literario y también

² <<https://humanidadesdigitaleshispanicas.es/>> (cons. 05/11/2021). Las actas se publicaron en López Poza y Pena Sueiro (2014).

aparecen en línea las primeras ediciones que incorporan lenguajes de marcado como el XML-TEI³.

La base de datos COMEDIC en las Humanidades Digitales españolas

La base de datos COMEDIC, acrónimo de «Catálogo de obras medievales impresas en castellano», nace en el año 2012, precisamente dentro de este contexto de mayor impulso de las Humanidades Digitales y de incremento del número de bases de datos. Fruto de la iniciativa de unos profesores del Departamento de Filología Española de la Universidad de Zaragoza pertenecientes al grupo de investigación consolidado CLARISEL, financiado por el Gobierno de Aragón, se desarrolla gracias a la concesión de un proyecto de investigación I+D+I dependiente del por aquel entonces Ministerio de Economía y Competitividad dentro del VI Plan Nacional de Investigación Científica, Desarrollo e Innovación⁴. El objetivo que perseguían era claro: la realización de un catálogo que recogiese todas las obras literarias medievales que han sobrevivido gracias a la imprenta, originales y/o traducidas, sin excluir los textos con una notable tradición manuscrita pero dejando a un lado la «literatura gris». El nexo de unión de estos investigadores era estudiar la relación entre la literatura medieval y la renacentista, concentrándose en la producción estrictamente castellana tanto dentro como fuera de la Península, lo que llevó a la incorporación de dos investigadores internacionales⁵. La premisa de la que se partió fue la no existencia de una ruptura entre la literatura medieval y la producida a partir de 1500⁶.

³ Resulta pionero el trabajo de edición digital de las *Soledades* de Góngora realizado por Rojas Castro (2015; 2017)

⁴ En el equipo inicial se encontraban María Jesús Lacarra, al frente del proyecto, Juan Manuel Cacho Blecua, Alberto del Río, María Carmen Marín Pina, José Aragüés Aldaz, María Sanz Julián y la doctoranda María Coduras Bruna.

⁵ Estos fueron Amaia Arizaleta (Université de Toulouse 2-Jean Jaurés) y Gaetano Lalomia (Università degli Studi di Catania). La primera se ocuparía de las obras en castellano impresas en Francia y el segundo de aquellas estampadas en territorio italiano.

⁶ Véanse al respecto las reflexiones que ya habían realizado Whinnom (1967) y Simón Díaz (1988).

Se sentaban así las bases para el inicio del trabajo, esto es, el estudio de la difusión, evolución, transformación y recepción de la literatura medieval española en el Quinientos, y la posibilidad de analizar la supremacía de algunos géneros literarios en detrimento de otros, así como las manipulaciones que sobre los textos operaron los tipógrafos para orientarlos a los nuevos lectores e introducirlos en los nuevos paradigmas de lectura (Santonocito, 2013). En relación con el *corpus*, cronológicamente se encuadró entre 1470, fecha en la que ve la luz la primera obra en castellano salida de una imprenta hispana, el *Sacramental* de Clemente Sánchez de Vercial, y 1600, momento en que se produjo el cambio hacia una recepción más moderna de los textos. El principal requisito que estos debían cumplir era haber conocido existencia previa a 1500.

El proyecto fue presentado ese mismo año en la *SEMYR* (Sociedad de Estudios Medievales y Renacentistas) y en el *Seminario de Estudios sobre Narrativa Caballeresca* de la Universidad Nacional Autónoma de México y, al año siguiente, en la *Asociación Hispánica de Literatura Medieval*. Los inicios de la creación de la base de datos fueron complejos, y se comenzó el trabajo sobre fichas creadas en *FileMaker* que contenían los principales campos que se incluirían, para después contar con una interfaz alojada en los servidores web de la Universidad de Zaragoza que permitiría hacer visibles los resultados al mismo tiempo que garantizaba a los investigadores seguir trabajando sobre las fichas. Además, el análisis de los textos no era solo textual o material, sino que implicaba un planteamiento multidisciplinar al prestar atención a las ilustraciones e imágenes, a su recepción y a los elementos paratextuales que acompañan a las imágenes, a lo que se sumaba la dificultad añadida de la lenta respuesta de algunas bibliotecas a la hora de proveer las digitalizaciones. Dos metas se perseguían: incrementar el número de fichas disponibles y aumentar la visibilidad de la base de datos (Lacarra, 2019, 421)⁷.

Para ello fue fundamental su integración en la red ARACNE en 2017,

⁷ El equipo del proyecto contaba ya con una experiencia previa en el uso de bases de datos y herramientas informáticas, puesto que desde finales de los 90 han sacado adelante tres bases de datos bibliográficas: *Amadís*, sobre literatura caballeresca, *Sendebarr*, sobre cuentística, y *Heredia*, sobre literatura aragonesa medieval. Están agrupadas en CLARISEL. Bases de datos bibliográficas, <<https://clarisel.unizar.es/>> (cons. 15/10/2021). Se suma la puesta en línea de DINAM. Diccionario de nombres del ciclo amadísiano, <<http://dinam.unizar.es/>> (cons. 15/10/2021).

la primera red de Humanidades Digitales y Letras Hispánicas en España y un hito en la investigación en Humanidades. Como iniciativa, ARACNE surge en julio de 2011 en el seno del *I Seminario Internacional sobre Bibliotecas Digitales y Bases de Datos especializadas para la investigación en Literaturas Hispánicas (BIDESLITE)* organizado por el Grupo de Estudios de Prosa hispánica bajomedieval y renacentista, dirigido por aquel entonces por Mercedes Fernández Valladares, con la colaboración del proyecto DIALOGYCA-BDDH y la ayuda del Instituto Universitario Menéndez Pidal⁸. El debate que se suscitó tomaba como punto de partida la dispersión que presentaban los proyectos de investigación en la aplicación de los conceptos y herramientas propuestos por las Humanidades Digitales puesto que, pese a tratarse de proyectos diversos en contenidos, objetivos, metodología, géneros literarios y periodos temporales, se habían enfrentado a problemas muy similares (Baranda Leturio y Rodríguez López, 2014, 101-102). La finalidad fue, entonces, intentar dar respuesta a la necesidad de coordinación y unificación de criterios percibida por los equipos de investigación de algunos proyectos de literatura española, precisamente ante una falta de unanimidad en la utilización de estas nuevas herramientas digitales (Pena Sueiro, 2017, 408)⁹. En 2012 consigue financiación bajo la dirección de Pedro Ruiz Pérez¹⁰ y se encamina hacia tres líneas diferentes: la consecución de una mayor visibilidad y difusión de los trabajos realizados, que esa visibilidad fuese un elemento de identificación y desarrollo de líneas de investigación con base tecnológica de calidad y la conveniencia de generar instrumentos de enlace y mejora en la recuperación de los datos y el desarrollo de herramientas que permitiesen proporcionar información de calidad para la creación de otras herramientas en el área (Baranda Leturio y Rodríguez López, 2014, 101).

El resultado ha sido la creación por parte del Laboratorio de Bases

⁸ Las actas de este encuentro están disponibles en: <<https://webs.ucm.es/BUCM/blogs/Foliocomplutense/4104.php>> (cons. 16/10/2021).

⁹ Los seis proyectos fundadores de ARACNE fueron BIESES (UNED), BSF (UDC), BIDISO (UDC), CLARISEL (UZ), DIALOGYCA (UCM) y PHEBO (UCO).

¹⁰ Ministerio de Economía y Competitividad (Gobierno de España), referencia de proyecto FFI2011-15606-E. El equipo inicial de trabajo estuvo formado por Pedro Ruiz e Ignacio García Aguilar (UCO), Nieves Baranda y María Dolores Martos (UNED), Ana Vian, Consolación Baranda y Mercedes Fernández Valladares (UCM), José Luis Villacañas Berlanga (UCM), Juan Manuel Cacho Blecua y María Jesús Lacarra y María Carmen Marín Pina (UZ), Sagrario López Poza y Nieves Pena Sueiro (UDC).

de Datos de la Universidad da Coruña de un portal que facilita el acceso a catorce recursos de investigación. Sobre este portal se ha diseñado un buscador de metadatos estándar Dublin Core que facilita una consulta simultánea de todos los contenidos ofrecidos por los distintos proyectos que lo componen gracias al protocolo OAIPMH (*Open Archive Initiative-Protocol for Metadata Harvesting*), que permite que el contenido de las colecciones contenidas en Aracne sea visible en recolectores virtuales como Hispana o Europeana (Arrigoni y Rodríguez López, 2014, 250-251)¹¹. La incorporación de la base de datos COMEDIC no fue sino una continuación por parte de estos investigadores de la línea seguida por sus bases de datos CLARISEL, en los orígenes de la red, y un paso más hacia la difusión de los contenidos investigadores. El punto culminante llegaría en 2019 con la concesión por parte del Ministerio de la Red de Excelencia ARACNE NODUS, bajo la dirección de Nieves Pena Sueiro, que pretende seguir con la búsqueda de una mayor visibilidad de los resultados de sus colecciones (Pena Sueiro, 2017, 411)¹².

Fruto de renovaciones posteriores del proyecto, la base de datos ha ido implementándose, así como ha aumentado progresivamente el número de obras actualmente visibles para los investigadores, que se cuenta en 83. La interfaz que se presenta al usuario por cada una de las fichas que la componen es sencilla: se recogen de manera general las variantes del título que presenta la obra en todas sus ediciones conservadas, así como los segundos autores, destinatarios o comisionarios, las fechas de redacción y/o traducción y los testimonios manuscritos conservados. Se da cuenta también de las ediciones modernas de la obra, los testimonios de lectura y, punto especialmente trascendental, la reescritura que sufre esta precisamente con motivo de su tradición impresa. Una sección destinada a la bibliografía cierra el conjunto. En el interior de las fichas se abre un indicador por cada edición documentada

¹¹ Sobre su descripción y funcionamiento, ver Alvite Díez y Pena Sueiro (2020).

¹² <<https://www.red-aracne.es/presentacion>> (cons. 11/11/2021). Ministerio de Ciencia, Innovación e Universidades, referencia de proyecto RED2018-102755-T. La concesión de la Red de Excelencia vino de la mano de la incorporación de cinco nuevos grupos, cuatro de perfil humanista y uno tecnológico. Estos son: BeCLar (dir. Antonio Moreno Hernández, UNED), PARNASEO (dir. Marta Haro Cortés, UV), Progetto Mambrino (dir. Stefano Neri, UNIVR), SILEM (dir. Carlos María Collantes Sánchez, US) y el LBD (dir. Ángeles Saavedra Places, UDC).

como fiable. De estas se refieren los repertorios en los que aparecen recogidas, tanto en formato papel como en formato digital; los ejemplares conservados, incluidos los digitalizados en bibliotecas digitales¹³; los facsímiles, en el caso en que los haya; y, finalmente, los paratextos entendidos según el concepto de Genette (2001): legales, socioliterarios, editoriales y grabados interiores y de portada¹⁴.

COMEDIC Búsqueda Catálogo Abreviaturas Grupo Clarisel Login

Ficha: CMDC91
Autor-es: San Pedro, Diego de (ca. 1450-post. 1498)
Título normalizado: *Arnalte y Lucenda*
Variantes título: *Tractado de amores de Arnalte a [sic] Lucenda* (Burgos, 1491, portada); *Tratado llamado Sant Pedro a las damas de la Reina nuestra señora* (Burgos, 1491, colofón; toma el título del encabezamiento del prólogo, no de la portada); *Arnalte y Lucenda* (Burgos, 1522, portada); *Tratado de Arnalte y Lucenda por elegante y muy gentil estilo, hecho por Diego de Sant Pedro y enderegado a las damas de la muy alta, católica y muy esciarscida reina doña Isabel. En el cual hallarán cartas y razonamientos de amores de mucho primor y gentileza según que por él verán* (1522, portada). "Aquí se acaba el libro de *Arnalte y Lucenda*"(1522, colofón).
Destinatario: "A las damas de la Reina" (1491, h. a2r). "San Pedro, criado del conde de Hureña, a las damas de la Reina nuestra señora" (1522, h. a1v).
Fecha composición: ca. 1481
Testimonios manuscritos: Se conservan dos manuscritos: Madrid. BNE, ms. 22021, fols. 13r-63r y Milán. ATM, ms. 940, fols. 133v-222v, que parecen ser copias de testimonios impresos. Ambos son importantes en el estudio textual, pues las ediciones impresas conservadas son muy defectuosas, con descuidos achacables al trabajo de imprenta y que derivan en lecturas ininteligibles. El ms. M (BNE) solventa algunas veces estos errores y parece presentar un texto superior a los impresos (vid. Alvar-Lucía. Repertorio, p. 395). El ms. T (Biblioteca Trivulziana) fue dado a conocer por Giovanni Caravaggi, "Un manuscrit espagnol inédit et un cas curieux de tradition textuelle", *Marche Romane*, 23-24 (1973-1974), pp. 157-168. A él se debe también la edición posterior del mismo en su *Miscellanea spagnola della "Trivulziana"*, Firenze, Leo S. Olschki, 1976. Lo valoran en relación con los impresos, Ivy A. Corfis, "Tractado de amores de Arnalte y Lucenda: Ms. 940 of the Biblioteca Trivulziana, Milán", *La corónica*, 14 (1985), pp. 36-39, y Mazzocchi (2004: 375; 2009: 203), que remite a la tesis de Martina Ricci, *La tradición del "Arnalte y Lucenda" de Diego de San Pedro. una nueva propuesta estemática*, *Tesi di laurea in Lingue e Letterature Straniere*, Ferrara, Università degli Studi di Ferrara, 2003-2004. El texto del códice trivulziano "más cercano al impreso de 1522 en sus lecturas, es una copia tardía en la que se han incorporado tanto en el léxico como en la ortografía italianismos propios de la naturaleza de la copia" (vid. Alvar-Lucía. Repertorio, p. 395).
Testimonios impresos: Cinco impresos: 1) Burgos: Fadrique Biel de Basilea, 1491, 25 de noviembre; 2) Burgos: Alonso de Melgar, 1522; 3) [Sevilla]: s.i., [1525]; 4) Burgos: [Viuda de Alonso de Melgar y Juan de Junta], 1527; 5) Valencia: Francisco Díaz Romano, 1535.

1): Burgos: Fadrique Biel de Basilea, 1491, 25 de noviembre

2): Burgos: Alonso de Melgar, 1522

3): [Sevilla]: s.i., [1525]

4): Burgos: [Viuda de Alonso de Melgar y Juan de Junta], 1527

5): Valencia: Francisco Díaz Romano, 1535

Fig. 1. COMEDIC. Ficha de *Arnalte y Lucenda*. Autora: María Carmen Marín Pina

Una última implementación se ha llevado a cabo, estrechamente relacionada con uno de los objetivos a corto plazo que la red ARACNE se estableció en su fundación: el «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

¹³ Está disponible también para los usuarios un manual donde se recogen las principales referencias abreviadas empleadas en los repertorios bibliográficos citados y las principales bibliotecas patrimoniales y digitales donde se encuentran depositados los ejemplares.

¹⁴ Véase Cacho Bleuca (2014) y Lacarra (2017).

reconocimiento y evaluación de los trabajos realizados» (Baranda y Rodríguez López, 2014, 106). La ausencia de criterios que permitan la evaluación de la calidad investigadora presente en la elaboración de materiales científicos dentro de las Humanidades Digitales por parte de las agencias estatales¹⁵ ha intentado solventarse en cada una de las entradas de la base de datos mediante la asignación de un DOI y un ISSN que las haga evaluables¹⁶.

Transkribus y COMEDIC: una propuesta de integración

Transkribus y los modelos de trabajo

Con la llegada del nuevo milenio se incrementó el número de tecnologías avanzadas destinadas al tratamiento de imágenes digitalizadas, triunfando los *softwares* de reconocimiento de caracteres, gracias a este aumento de la financiación desde los gobiernos nacionales y autonómicos, los proyectos a nivel internacional y europeo y la intervención de instituciones y fundaciones privadas. Los OCR (*Optical Character Recognition*) se definen como sistemas de conversión de *inputs* en forma de texto en un formato codificado por la máquina. Frecuentes en *softwares* comerciales u *Open Source*, dependen principalmente de la extracción de rasgos y la discriminación y/o clasificación de estos rasgos basados en una serie de patrones (Memon, Sali *et al.*, 2020). Los sistemas de reconocimiento de caracteres son herramientas específicamente creadas para la obtención de transcripciones a partir de imágenes escaneadas de las fuentes en papel, y aplicados sobre una imagen de una página impresa, proceden a fragmentarla en caracteres que se comparan con conjuntos de

¹⁵ «Igualmente parece poco razonable que quien pone al servicio de la comunidad el fruto de su trabajo publicándolo en acceso abierto enseguida (lo cual puede ayudar a acelerar otras investigaciones) sea menos valorado que el que publica un libro cuya distribución es deficiente. Es chocante —y delata ignorancia— que, en las evaluaciones de producción científica, se siga pidiendo como condición que una publicación lleve un número de ISBN, como si ello fuera un marchamo de calidad (cuando no es más que un control de ventas que cualquiera puede solicitar a la Federación de Gremios de Editores de España) y no se midan otros parámetros más científicos» (López Poza, 2015, 5).

¹⁶ Lucía Megías (2010, 1253) aboga precisamente por acercarse al modelo de patentes propuesto por las disciplinas científicas pero adaptado al contexto de las Humanidades.

rasgos abstractos que describen caracteres prototípicos aprendidos previamente a partir de un entrenamiento sobre las tipografías. Este se basa en metodologías de Aprendizaje Automático (*Machine Learning*) y Aprendizaje Profundo (*Deep Learning*) que se aplican primero sobre material preparado y anotado. Se trata de un proceso que funciona especialmente bien con tipografías modernas, puesto que las similitudes entre las fuentes aprendidas y reconocidas, la clara separación entre caracteres uniformes sobre un fondo blanco impoluto y una ortografía moderna y estandarizada contribuyen al buen reconocimiento (Springman y Lüdeling, 2017). Su empleo reciente sobre caligrafía manuscrita ha recibido bastante atención, especialmente por su contribución a la digitalización de manuscritos medievales (Memon, Sali *et al.*, 2020).

Con todo, los sistemas OCR no han dejado de estar exentos de riesgos y problemáticas. Las manchas en las páginas de los testimonios, debidas a la humedad o al deterioro por el paso del tiempo, sumadas a las deformaciones procedentes del escaneo manual o del estado defectuoso de las fuentes empleadas han dificultado la aplicación del reconocimiento sobre ciertos soportes (Springman y Lüdeling, 2017). Esto afecta directamente al tratamiento de textos anteriores a la época contemporánea, especialmente a las obras impresas en el siglo XVI, centuria en la que se inscribe el corpus de obras trabajado en COMEDIC, donde convivían diferentes tipografías (gótica, redonda, cursiva) que podían adquirir incluso variaciones internas dependiendo de los distintos juegos de tipos y fundiciones de cada impresor¹⁷. Se suman otras dificultades como el uso frecuente de abreviaturas, el empleo de signos tironianos, las ligaduras entre caracteres y factores externos como los desgastes de los tipos y los defectos de entintado (Bazzaco, 2020, 545). Además, las tecnologías OCR imponen una correspondencia directa entre el signo ortográfico y una letra en el texto transcrito, lo que obliga a entrenar un único modelo individual para un libro individual y una tipografía concreta. Este mecanismo, que podía funcionar para el libro o texto que había servido de entrenamiento, se volvía impracticable al aplicarlo a otros diferentes puesto que, aunque la fuente resultase similar

¹⁷ Al respecto, ver también Barbuti y Caldarola (2013) y Berk-Kirpatrick y Klein (2014).

para el ojo humano, los resultados arrojados no eran positivos (Sprigmann y Lüdeling, 2017). Para paliar este defecto, se han implementado las plataformas de HTR (*Handwritten Text Recognition*), donde se integra Transkribus, nacidas para la lectura e interpretación de textos manuscritos, pero aplicables también a textos impresos de esta primera época de implantación de la imprenta en nuestro país. Esta plataforma, surgida del Proyecto READ (*Recognition and Enrichment of Archival Documents*) dentro del programa europeo Horizon 2020¹⁸, se basa en la colaboración entre usuarios que proveen los materiales digitalizados para generar *inputs* mientras un equipo de informáticos se encargan del procesamiento de los datos. Dichos materiales han tenido que pasar por un tratamiento previo de la imagen para asegurar su calidad y garantizar el correcto reconocimiento.

El trabajo que pretende realizarse en COMEDIC se integra en los avances realizados por el Progetto Mambrino en la aplicación de Transkribus. Con el objetivo de elaborar ediciones digitales académicas de las obras que componen el ciclo italiano del *Amadís de Gaula*, comenzaron los primeros proyectos experimentales en el año 2016 con la tipografía cursiva que caracteriza a estos impresos (Mancinelli, 2016; Bazzaco, 2018). El objetivo ahora es la elaboración de un modelo de reconocimiento o *extended model* de letra gótica que sea aplicable a cualquier texto impreso con esta tipografía a lo largo del Quinientos. La creación de este modelo permitiría a los usuarios en un futuro transcribir cualquier tipo de texto en gótica de este periodo con un reducido margen de error a la hora de lanzar el reconocimiento automático. Para ello, se partió de la selección de un pequeño corpus de textos que presentasen cierta homogeneidad en su conformación tipográfica, cierta extensión y que poseyesen estándares de calidad en su digitalización en formato imagen. Además, como rasgo definitorio, estos debían pertenecer a impresores y lugares diferentes, estar estampados en tipografía gótica y datarse en un periodo temporal amplio que finalmente ha abarcado desde 1526 hasta 1563. El listado de obras con el que empezaron a trabajar fue el siguiente (Bazzaco, 2020, 549):

¹⁸ <<https://readcoop.eu/>> (cons. 8/10/2021). Se celebran encuentros anuales donde investigadores ponen en común sus resultados en el empleo de la herramienta, ver por ejemplo el celebrado en 2020: <<https://readcoop.eu/transkribus-user-conference-2020>> (cons. 29/10/2021).

Título	Autor	Impresor(es)	Localización
<i>Lisuarte de Grecia</i>	Juan Díaz	Jacobo y Juan Cromberger Sevilla, 1526	BNE R/71
<i>Florando de Inglaterra</i>	Anónimo	German Gallarde Lisboa, 1545	BL C.62.h.14.
<i>Silves de la Selva</i>	Pedro de Luján	Dominico de Robertis Sevilla, 1549	BNE R/865
<i>Leandro el Bel</i>	Pedro de Luján	Miguel Ferrer Toledo, 1563	BNE R/9030

Tabla 1. Obras que constituyen el *dataset* del trabajo realizado por el Progetto Mambrino

De esta manera, se inició el proceso de aprendizaje automático o *Machine Learning* (Carbonell, Michalski *et al.*, 2013) mediante la creación de unos algoritmos que sirvieran de base (o *groundtruth*) que buscasen una correlación entre los signos presentes en la imagen y sus respectivas transcripciones que permitiese llegar a la creación de dicho modelo extendido (Jander, 2016). Los resultados iniciales arrojaron buenos augurios, lo que llevó a la necesidad de plantear una aproximación colaborativa y un enriquecimiento de los materiales de base usados para el entrenamiento, pues solo mediante el incremento de transcripciones posteriores podía llegarse a implementar un reconocimiento más fiable de los textos sobre los que se aplicaba el modelo (Bazzaco, 2020, 552).

Es en esta etapa donde COMEDIC se suma a la iniciativa, añadiendo para la creación del modelo extendido o *extended model* de gótica los siguientes títulos a modo de textos de entrada (Tabla 2).

La peculiaridad de los textos propuestos por nuestra base de datos, además de mostrar una gran variedad de tipos pertenecientes a diferentes impresores del Quinientos, radica fundamentalmente en una *dispositio textus* a línea tirada, frente a la doble columna de los textos caballerescos sugeridos por Progetto Mambrino para el modelo de gótica. Se añaden además otras particularidades de puesta en página, como las que presentan la *Tragicomedia de Calisto y Melibea* (Roma, Marcellus Silber, s.a.) o el *Retablo de la vida de Cristo* de Juan de Padilla (Sevilla, Juan Cromberger, 1510)¹⁹, con el objetivo de enriquecer el *Layout Analysis* (análisis de la *mise en page*).

¹⁹ Remito para ello a la contribución de Bazzaco *et al.* (pp. 67-125) en este mismo volumen.

Título	Autor	Lugar de edición y fecha
<i>Historia de la linda Magalona</i>	Anónimo	Sevilla, Jacobo Cromberger, 1519
<i>Historia de la reina Sebilla</i>	Anónimo	Burgos, Felipe de Junta, 1551
<i>Historia del rey Canamor</i>	Anónimo	Valencia, Jorge Costilla, 1527
<i>Libro del conde Partinuplés</i>	Anónimo	Sevilla, Jacobo Cromberger, 1519
<i>Libro del conde Partinuplés</i>	Anónimo	Burgos, Herederos de Juan de Junta 1558
<i>Libro del conde Partinuplés</i>	Anónimo	Burgos, Felipe de Junta, 1563
<i>Doctrinal de los Caballeros</i>	Alonso de Cartagena	Burgos, Fadrique Biel de Basilea, 1487
<i>La Fiameta</i>	Juan Boccaccio	Salamanca, [Impresor de la Gramática de Nebrija], 1497
<i>Crónica del Rey Don Rodrigo (Crónica Sarracina)</i>	Pedro de Corral	[Sevilla], [Meinaro Ungut y Estanislao Polono], 1499
<i>Tragicomedia de Calisto y Melibea</i>	Fernando de Rojas	Roma, Marcellus Silber, a. 1515
<i>Retablo de la Vida de Cristo</i>	Juan de Padilla	Sevilla, Jacobo Cromberger, 1505

Tabla 2. Obras que constituyen el *dataset* del trabajo realizado por los colaboradores de COMEDIC

En relación con los criterios de transcripción, la propuesta de futuro pasaba por trabajar sobre dos modelos distintos de reconocimiento, uno basado en transcripciones más conservadoras, y otro basado en criterios de transcripción más basado en criterios de transcripción con un mayor grado de modernización (Bazzaco, 2020, 561). Finalmente, nos hemos inclinado por el primero, que facilita un patrón de reconocimiento con un índice de error algo más reducido. Sin embargo, se ha procedido a desarrollar las abreviaturas, que se han indicado mediante un sistema de etiquetado que proporciona la propia plataforma Transkribus. El resultado final ha sido presentado en el congreso *Humanidades Digitales y estudios literarios hispánicos. De los impresos de la Edad Moderna a las ediciones académicas digitales*, celebrado en Verona del 22 al 23 de junio de 2021 bajo la coordinación académica de Stefano Bazzaco.

La integración de Transkribus en el trabajo de COMEDIC

Actualmente nos situamos en lo que ha venido a denominarse como «segunda textualidad», coincidente con el momento en que la escritura se transforma tras surgir en Grecia en el siglo VIII a.C tal y como la conocemos en la cultura occidental, que permite conjugar características de la primera textualidad con algunas características de la oralidad y la introducción del nuevo soporte traído por la informática. Su modelo representativo será el texto digital (Lucía Megías, 2012b, 263-165), definido como un texto en cuyo proceso de difusión interviene una codificación de la información por parte de lenguajes artificiales, y que se presenta materialmente como información lingüística codificada matemáticamente y representada bajo una forma legible, interpretable y reconocible por el ser humano. Por ello, en su composición confluyen dos capas de información: una capa humana continuadora de la tecnología de la escritura, y que supone codificar y descodificar de manera lógica una serie de signos, y una capa de información matemática que entraña el procedimiento que transforma esa escritura en algo legible por la tecnología informática, el conocido como código binario, que le permite nuevos usos y nuevos modelos de relación (Paisão de Sousa, 2009).

A la hora de tratar la concepción del llamado texto digital, que incluye a su vez las posibilidades que ofrece la Web 2.0, entre ellas la interacción con los propios receptores, son imprescindibles los conceptos de «digitalización» o «texto digitalizado», que supone el traslado de una imagen con texto escrito como imagen fija al medio digital y «digitación» o «texto digitado», que incluye necesariamente la intervención de la codificación estándar ASCII y Unicode propia de un procesador (Bazzaco, 2020, 537). Lucía Megías (2012a, 115-116) ha establecido una división en tres estadios distintos que permiten llegar a este concepto y que suponen una gradación entre los tres aspectos de la digitalización textual teniendo en cuenta la finalidad, la tecnología y la relación con los medios de transmisión analógica, en la que los conceptos anteriores entroncan a la perfección:

- 1) La reproducción digital de un manuscrito o un impreso por medio de la fotografía digital y el escaneado, la cual supone una fase de

acumulación de información que desarrollan, principalmente, las bibliotecas digitales virtuales. El resultado es un objeto textual «digitalizado».

2) La creación o digitalización de textos con la pretensión de ser difundidos fuera del ambiente y los medios de transmisión digitales, especialmente en el medio impreso. Son los textos generados por los procesadores de textos que cierran el texto en una única imagen y que han sufrido un proceso de «digitación».

3) Texto digital que utiliza procesos de codificación pensados para poder ser visualizados en la pantalla del ordenador aprovechando la hipertextualidad y la posibilidad de establecer relaciones entre el nivel estructural y semántico, y para el que cobran especial importancia los lenguajes de marcado HTML, XML, XHTML o XML-TEI. En esta etapa no se intenta emular en el medio digital los modelos textuales del mundo analógico, sino que lo que se pretende es profundizar en modelos que ofrezcan experiencias variadas a la unión del creador, del lector y del propio medio que ofrece este modelo.

Actualmente el proyecto se encuentra conjugando las dos primeras fases. La recopilación en los indicadores de las fichas de los ejemplares de cada edición que se encuentran digitalizados permite esa acumulación de información que facilita el análisis de este objeto libro casi analógico para proceder a su descripción, al mismo tiempo que ahorra desplazamientos de los investigadores, algo que se han convertido en indispensable desgraciadamente en los tiempos actuales. Mediante el campo «ediciones modernas», consignamos esas ediciones en las que el texto ha pasado un proceso de «digitación» para ver luego la luz de nuevo en formato analógico, si bien cada vez más se ofrece la posibilidad de su consulta en línea en formatos como epub, el propio de los libros electrónicos. El análisis de las ediciones de obras concretas también está propiciando la creación de nuevas ediciones críticas mediante tesis doctorales que van a ver la luz de nuevo también en formato cercano al analógico, si bien los repositorios digitales facilitan su consulta en la red. Todo ello inserta a COMEDIC dentro de las llamadas «bibliotecas de investigación» —«aquellas que intentan aunar el rigor filológico con el temático agrupando normalmente textos de una o varias áreas de conocimiento»

(Canet, 2005, 151)—, pertenecientes al grupo de las que solo incorporan las reproducciones facsímiles y cuyo trabajo está encabezado por un profesor universitario.

Mediante la incorporación de Transkribus pretendemos dar un paso adelante e integrarnos en un grado intermedio entre las fases 2 y 3. En primer lugar, nos centraremos en el reconocimiento de obras de extensión breve que posean un número reducido de ediciones, tomando como punto de referencia la *princeps*, o que tan solo se hallen conservadas en una. Esta transcripción será paleográfica, de acuerdo con los modelos de entrenamiento que en la actualidad hemos desarrollado para el modelo de HTR *SpanishGothic_XV-XVI_extended*, y figurará mediante un hipervínculo en el apartado «ediciones modernas». El hipervínculo conducirá a la transcripción, alojada en la nube, en dos formatos, el pdf en la fase 2 y HTML, ya en la fase 3, mediante la introducción de un lenguaje de marcado capaz de ser reconocido y visualizado por un ordenador y que permite abrir nuevas vías dentro de la edición digital. COMEDIC evolucionaría así hacia una nueva configuración de la «biblioteca de investigación» a través de la publicación del texto y el facsímil, y se gestaría como una hibridación entre una base de datos y una biblioteca digital «que no solo permite realizar búsquedas complejas, sino que también da acceso a textos codificados en HTML y/o formato pdf» (Rojas Castro, 2013, 43). De acuerdo con los criterios que ya rigen la base, cada edición contará con un responsable y un revisor que examine el resultado.

Nuevas vías de futuro que se abren

Editar en la actualidad supone ser consciente del cambio que la sociedad y el mundo académico está viviendo dentro del paradigma digital, y esto implica que el editor digital debe ahora conocer y dominar gran parte de las herramientas informáticas que tiene a su disposición para fijar el texto de manera científica y difundirlo en el nuevo formato (Lucía Megías, 2019, 98). Actualmente nos encontramos en el «incunable del hipertexto», concepto con el que Lucía Megías (2007) pretende establecer una analogía entre estos momentos iniciales del texto digital y los

momentos iniciales del texto impreso. Mostrar una transcripción paleográfica del texto de una obra que no ha sido editada o que no cuenta con ediciones modernas puede servir de punto de partida para elaborar nuevas ediciones críticas analógicas mediante la adaptación del texto a los criterios de edición que el editor quiera aplicar. Además, puede contribuir a la legibilidad de documentos que se han visto dañados o cuya tinta se ha desdibujado, teniendo en cuenta siempre que hay que prestar atención a informaciones a las que no se podrá acceder, como las características físicas del documento, la calidad del papel o las marcas de agua (Ogilvie, 2017, 82).

No obstante, la posibilidad de ofrecer nuestros textos transcritos en lenguaje de marcado HTML abre nuevas vías que encaminan nuestro trabajo a la consecución de los objetivos del texto digital. En primer lugar, facilita la comparación entre la imagen digitalizada y el texto transcrito, pero además permite implementar esta transcripción aplicándole un sistema de marcado semántico XML-TEI potenciando sus posibilidades en la red. Esto permitiría también avanzar hacia la posible creación de una edición sinóptica integral, aquella que permite poder acceder simultáneamente a las transcripciones de todos los testimonios conservados de una obra, así como a sus variantes a partir de la lección de uno de ellos. En segundo lugar, facilita asentar los cimientos que abran el camino a la edición crítica hipertextual. Esta parte de la presentación analógica, pero se sirve del nuevo medio de transmisión al mismo tiempo que conserva el modo habitual de trabajar del filólogo. Mediante el añadido de la hipertextualidad y la hipermedialidad²⁰, se permitiría el acceso a la obra desde varias aproximaciones: la paleográfica, el análisis lingüístico o el estudio de la evolución textual. Supondría el establecimiento de una relación emergente de transformación donde el texto empieza a modificarse para adaptarse a las características y posibilidades del nuevo medio conjugando las tres morfologías de la información (texto, imagen y sonido) (Lucía Megías, 2009, 12). De esta forma, desembocaría en un entramado hipertextual que pondría en

²⁰ La hipertextualidad, la hipermedialidad y la intertextualidad han sido definidos como los tres principales rasgos definitorios del ámbito digital y, por lo tanto, deberían estar presentes en una edición digital.

relación las digitalizaciones de los facsímiles conservados, las transcripciones paleográficas, las ediciones críticas del texto, la bibliografía y cualquier información adicional que ofreciera posibilidades de acercamiento a la obra (Lucía Megías, 2019, 109).

Sea como fuere, el texto siempre estará configurado como el elemento central sobre el que se podrá acceder al resto de materiales, que tendrán que estar estrechamente vinculados aprovechando las herramientas informáticas que han aparecido en los últimos años. COMEDIC pretende sembrar la semilla para que una de estas nuevas realidades de estudio y análisis del objeto literario sea posible.



Bibliografía citada

- Alvite Díez, María Luisa y Nieves Pena Sueiro, «Colecciones digitales patrimoniales especializadas. Estudio de la Red ARACNE», en *Actas del IV Congreso ISKO España-Portugal 2019. XIV Congreso ISKO España (Barcelona, 11 y 12 de julio de 2019)*, eds. Jesús Tramullas, Piedad Garrido-Picazo y Gonzalo Marco Cuenca, Zaragoza, Sociedad Internacional para la Organización del Conocimiento (ISKO)-Capítulo Ibérico, 2020, pp. 185-195.
- Arrigoni, Eleonora y Eduardo Rodríguez López, «La red de investigación de “Humanidades Digitales y Letras Hispánicas”: avance de la Red-Aracne», en *Visibilidad y divulgación de la investigación desde las Humanidades Digitales: experiencias y proyectos*, coord. Álvaro Baraibar Echeverría, Pamplona, Universidad de Navarra/GRISO (Grupo de Investigación Siglo de Oro), 2014, pp. 243-251. URL: <<https://dadun.unav.edu/handle/10171/35723>> (cons. 28/10/2021).

- Baranda Leturio, Consolación y Eduardo Rodríguez López, «Red ARACNE: retos y objetivos de un proyecto de coordinación en letras hispánicas digitales», en *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, eds. Sagrario López Poza y Nieves Pena Sueiro, *Janus*, Anexo 1 (2014), pp. 101-109. URL: <<https://www.janusdigital.es/anexos/contribucion.htm?id=10>> (cons. 28/10/2021).
- Barbuti Nicola y Tommaso Caldarola, «An Innovative Character Recognition for Ancient Book and Archival Materials: A Segmentation and Self-learning Based Approach», en *Digital Libraries and Archives. IRCDL 2012. Communications in Computer and Information Science*, eds. Maristella Agosti, Floriana Esposito, Stefano Ferilli y Nicola Ferro, Berlin-Heidelberg, Springer-Verlag, Berlin, Heidelberg, 2013, pp. 261-270. DOI: <https://doi.org/10.1007/978-3-642-35834-0_26> (cons. 28/10/2021).
- Bazzaco, Stefano, «El Proyecto Mambrino y las tecnologías OCR: estado de la cuestión», *Historias fingidas*, 6 (2018), pp. 257-272. DOI: <<https://doi.org/10.13136/2284-2667/89>> (cons. 15/10/2021).
- , «El reconocimiento automático de textos en letra gótica del Siglo de Oro: creación de un modelo HTR basado en libros de caballerías del siglo XVI en la plataforma Transkribus», *Janus. Estudios sobre el Siglo de Oro*, 9 (2020), pp. 534-561. URL: <<https://www.janusdigital.es/articulo.htm?id=160>> (cons. 15/10/2021).
- Berg-Kirkpatrick, Taylor y Dan Klein, «Improved Typesetting Models for Historical OCR», en *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Baltimore-Maryland, Association for Computational Linguistics, 2014, pp. 118-123. URL: <<https://aclanthology.org/P14-2020.pdf>> (cons. 28/10/2021).
- Cacho Blecua, Juan Manuel, «Hacia un catálogo de los textos medievales impresos (COMEDIC): el ejemplo de la *Crónica popular del Cid*», en *Texto, edición y público lector en los albores de la imprenta*, eds. Marta Haro Cortés y José Luis Canet, València, PUV, 2014, pp. 29-52.
- Canet, José Luis, «Bibliotecas digitales españolas a texto completo», *Syntagma. Revista del Instituto de Historia del Libro y de la Lectura*, 1 (2005), pp. 149-159.

- , «Reflexiones sobre las Humanidades Digitales», en *Janus*, Anexo 1: *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, eds. Sagrario López Poza y Nieves Pena Sueiro, (2014), pp. 11-20. URL: <<https://www.janusdigital.es/anexos/contribucion.htm?id=4>> (cons. 15/10/2021).
- Carbonell, Jamie G, Ryszard S. Michalski, Tom M. Mitchell, «An overview of machine learning», en *Machine Learning: An Artificial Intelligence Approach*, eds. Jamie G. Carbonell, Ryszard S. Michalski y Tom M. Mitchell, Berlin-Heidelberg, Springer-Verlag, 2013, pp. 3-23.
- Faulhaber, Charles, y Francisco Marcos Marín, «La conservación y utilización de textos en el futuro inmediato: ADMYTE, el archivo digital de manuscritos y textos españoles», *Hispania* 75/4 (1992), pp. 1010-1023.
- Genette, Gérard, *Umbrales*, Madrid, Siglo XXI, 2001.
- González-Blanco García, Elena, «Actualidad de las Humanidades Digitales y un ejemplo de ensamblaje poético en la red», *Cuadernos hispanoamericanos*, 761 (2013), pp. 53-67.
- Hernández Lorenzo, Laura, «Humanidades Digitales y Literatura española: 50 años de repaso histórico y panorámica de proyectos representativos», *Janus. Estudios sobre el Siglo de Oro*, 9 (2020), pp. 562-595. URL: <<https://www.janusdigital.es/articulo.htm?id=154>> (cons. 04/11/2021).
- Jander, Melina, «Handwritten Text Recognition–Transkribus: A User Report», en *The electronic Text Reuse Acquisition Project (eTRAP)*, Göttingen, Institute of Computer Science, University of Göttingen, 2016. URL: <<http://www.etrp.eu/transkribus-a-user-report/>> (cons. 05/11/2021).
- Karlsson, Lina, y Linda Malm, «Revolution or Remediation? A Study of Electronic Scholarly Editions on the Web», *HUMANIT*, 7/1 (2004), pp. 1-46. URL: <<https://humanit.hb.se/article/view/135>> (cons. 11/11/2021).
- Lacarra, María Jesús, «COMEDIC: un “Catálogo de obras medievales impresas en castellano” en construcción», en *En Doiro antr'o Porto e Gaia. Estudos de Literatura Medieval Ibérica*, ed. J. C. Ribeiro Miranda, Porto, Estratégias criativas, 2017, pp. 599-610.

- , «Comedic», *Historias fingidas*, 7 (2019), pp. 419-42. DOI: < <https://doi.org/10.13136/2284-2667/137> > (cons. 18/10/2021).
- López Poza, Sagrario, «Humanidades Digitales y literaturas hispánicas: presente y futuro», *Ínsula. Revista de letras y ciencias humanas*, 833 (2015), pp. 3-5.
- , «Humanistas y Humanidades Digitales. Trayectoria y proyección en la Filología española», en *Humanidades y humanismo. Homenaje a María Pilar Cuartero*, eds. Aurora Egido, José Enrique Laplana y Luis Sánchez Laílla, Zaragoza, Institución Fernando el Católico, 2019, pp. 125-160.
- López Poza, Sagrario y Nieves Pena Sueiro (eds.), *Janus*, Anexo 1: *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, (2014). URL: < <https://www.janusdigital.es/anexo.htm?id=5> > (cons. 03/11/2021).
- Lucía Megías, José Manuel, «La edición crítica hipertextual. Hacia la superación del incunable del hipertexto», en *Lecturas y textos en el siglo XXI. Nuevos caminos en la edición textual*, coords. Cristina Castillo Martínez y José Luis Ramírez Luengo, Lugo, Editorial Axac, 2009, pp. 11-74.
- , «Los nuevos filólogos del siglo XXI: La literatura medieval hispánica en la Red», en *Actas del XIII Congreso de la Asociación Hispánica de Literatura Medieval. In memoriam Alan Deyermond (Valladolid, 15 a 19 de septiembre de 2009)*, eds. José Manuel Fradejas Rueda, Déborah Dietrick Smithbauer, Demetrio Martín Sanz, María Jesús Díez Garretas, Valladolid, Universidad de Valladolid, 2010, pp. 1233-1254.
- , *Elogio del texto digital. Claves para entender el nuevo paradigma*, Madrid, Fórcola Ediciones, 2012a.
- , «El escritor en la era digital (un elogio a la segunda textualidad)», *Revista de Humanidades y Ciencias Sociales*, 2 (2012b), 255-269. URL: < http://rabida.uhu.es/dspace/bitstream/handle/10272/6332/El_escritor_en_la_era_digital.pdf?sequence=2 > (cons. 11/11/2021).

- , «El editor de textos ante el reto digital: elogio de la edición 2.0», *Revista de Humanidades Digitales*, 4 (2019), 93-114. URL: <<http://revistas.uned.es/index.php/RHD/article/view/25188>> (cons. 13/11/2021).
- Mancinelli, Tiziana, «Early printed edition and OCR techniques: what is the state-of-art? Strategies to be developed from the working-progress Mambrino project work», *Historias Fingidas*, 4 (2016), pp. 255-260. DOI: <<https://doi.org/10.13136/2284-2667/65>> (cons. 13/05/2022).
- Memon, Jamshed, Maira Samid, Rizwan Ahmed Khan y Mueen Uddin, «Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)», *IEEE Access*, 8 (2020). URL: <<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9151144>> (cons. 05/11/2021).
- Morràs, María y Antonio Rojas Castro (coords.), *Humanidades Digitales y Literaturas hispánicas*, *Ínsula*, 822 (2015).
- Ogilvie, Brian, «Scientific Archives in the Age of Digitization», *Isis*, 107/1 (2016), pp. 77-85. URL: <<https://www.journals.uchicago.edu/doi/full/10.1086/686075>> (cons. 03/11/2021).
- Paisão de Sousa, Maria Clara, «Conceito material de “Texto digital”: um ensaio», *Revista Texto Digital*, 5/2 (2009), pp. 159-187. URL: <<http://www.textodigital.ufsc.br/>> (cons. 11/11/2021).
- Pena Sueiro, Nieves y Ángeles Saavedra Places, «Aracne. Red de Humanidades Digitales y Letras Hispánicas», *Historias Fingidas*, 7 (2019), pp. 407-412. DOI: <<https://doi.org/10.13136/2284-2667/149>> (cons. 03/11/2021).
- Rojas Castro, Antonio, «El mapa y el territorio: una aproximación histórico-bibliográfica a la emergencia de las Humanidades Digitales en España», *Caracteres. Estudios culturales y críticos de la esfera digital*, 2/2 (2013), pp. 10-53.
- , *Editar las soledades de Góngora en la era digital: texto crítico y propuesta de codificación XML-TEI*, tesis doctoral, Barcelona, Universitat Pompeu Fabra, 2015.

- , «La edición crítica digital y la codificación TEI. Preliminares para una nueva edición de las *Soledades* de Luis de Góngora», *Revista de Humanidades Digitales*, 1 (2017), pp. 4-19. URL: <<http://revistas.une.es/index.php/RHD/article/view/16379>> (cons. 03/11/2021).
- Santonocito, Daniela, «Reescrituras y relecturas: hacia un catalogo de obras medievales impresas en castellano hasta 1600 (COMEDIC)», *Le forme e la storia*, VI/1 (2013), pp. 175-187.
- Simón Díaz, José, «La literatura medieval castellana y sus ediciones españolas de 1501 a 1560», en *El libro antiguo español. Actas del Primer Coloquio Internacional (Madrid, 18 al 20 de diciembre de 1986)*, eds. María Luisa López-Vidriero y Pedro M. Cátedra, Universidad de Salamanca/Biblioteca Nacional de Madrid/Sociedad Española de Historia del Libro, Salamanca, 1988, pp. 371-396.
- Spence, Paul, «Centros y fronteras: el panorama internacional», en Sagrario López Poza y Nieves Pena Sueiro (eds.), *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, *Janus*, Anexo 1 (2014), pp. 37-61. URL: <<https://www.janusdigital.es/anexos/contribucion.htm?id=6>> (cons. 03/11/2021).
- Spence, Paul y Elena González-Blanco García, «A Historical Perspective to Digital Humanities in Spain», *H-Soz-Kult* (2014). URL: <<https://www.hsozkult.de/debate/id/diskussionen-2449>> (cons. 04/11/2021).
- Springmann, Uwe y Anke Lüdeling, «OCR of historical printings with an application to building diachronic corpora», *Digital Humanities Quarterly*, 7/2 (2017). URL: <<http://www.digitalhumanities.org/dhq/vol/11/2/000288/000288.html#>> (cons. 24/10/2021).
- Toscano, Maurizio, Aroa Rabadán, Salvador Ros y Elena González-Blanco García (2020), «Digital humanities in Spain: Historical perspective and current scenario», *Profesional de la Información*, 29/6 (2020).
- Whinnom, Keith, «Spanish Literary Historiography: Three Forms of Distortion», en *An Inaugural Lecture Delivered in the University of Exeter on 8 December 1967*, 1967.