



PHILOLOGIA HISPALENSIS

ESTUDIOS LITERARIOS

2024 | VOL. XXXVIII **2**

PHILOLOGIA HISPALENSIS

AÑO 2024
VOL. XXXVIII/2

ESTUDIOS LITERARIOS



FACULTAD DE FILOLOGÍA
UNIVERSIDAD DE SEVILLA

EVALUACIÓN DE ORIGINALES: Los originales se someten a una evaluación ciega, un proceso anónimo de revisión por pares, siendo enviados a evaluadores externos y también examinados por los miembros del Consejo de Redacción y/o los especialistas del Consejo Asesor de la Revista.

PERIODICIDAD: Anual en formato tradicional y en formato electrónico.

PUBLICACIÓN EN INTERNET: <<https://editorial.us.es/es/revistas/philologia-hispalensis>>, <<https://revistascientificas.us.es/index.php/PH>>.

BASES DE DATOS: *Philologia Hispalensis* se encuentra indexada en CARHUS Plus+2018, CIRC (grupo B), DIALNET, DOAJ, Dulcinea, Index Islamicus, Latindex 2.0 (100% de los criterios cumplidos), MIAR (ICDS 2022 = 10), MLA, REDIB, SCOPUS, ERIHPLUS y ANVUR (Clase A). Asimismo, cuenta con el sello de calidad de la FECYT (8ª edición, 2023, renovado en 2024 y válido hasta 2025) en los campos de conocimiento Lingüística y Literatura dentro de la modalidad Humanidades.

ENVÍO DE ORIGINALES Y SUSCRIPCIONES: Las colaboraciones deben enviarse a través de <<https://revistascientificas.us.es/index.php/PH>>.

DIRECCIÓN DE CONTACTO: Secretariado de la Revista Philologia Hispalensis, Facultad de Filología, Universidad de Sevilla, C/ Palos de la Frontera, s/n, 41004 Sevilla; o bien al correo electrónico <philhisp@us.es>.

INTERCAMBIOS O CANJES (BIBLIOTECAS UNIVERSIARIAS): Solicítense a Editorial Universidad de Sevilla o al Secretariado de la revista <philhisp@us.es>.

© De los autores y Editorial Universidad de Sevilla

Financiación: Revista financiada por la Universidad de Sevilla dentro de las ayudas del VII PPIT-US y del Decanato de la Facultad de Filología.

PORTADA: referencias.maquetacion@gmail.com

DEPÓSITO LEGAL: SE-354-1986

ISSN: 1132 - 0265 / eISSN 2253-8321

Maquetación: referencias.maquetacion@gmail.com

IMPRIME: Podiprint

DISTRIBUYE: Editorial Universidad de Sevilla, Porvenir, 27, 41013 Sevilla

Licence Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0 DEED)



FECYT 641/2024
Fecha de certificación: 28 de julio de 2023 (9ª convocatoria)
Válido hasta: 24 de julio de 2025



EQUIPO EDITORIAL

Directora: Yolanda Congosto Martín, Universidad de Sevilla, España
Secretaria: Leyre Martín Aizpuru, Universidad de Sevilla, España
Editoras: Salomé Lora Bravo, Universidad de Sevilla, España
Natalia Silva López, Fundación Pública de Estudios Universitarios “Francisco Maldonado” de Osuna, España

Coordinadora de Reseñas: M. Amparo Soler Bonafont, Universidad Complutense de Madrid, España

Consejo de

Redacción: Gema Areta Marigó, Universidad de Sevilla, España
Elisabetta Carpitelli, Université Stendhal - Grenoble Alpes, France
Antonio Luis Chaves Reino, Universidad de Sevilla, España
Marianna Chodorowska-Pilch, University of Southern California, USA
Yves Citton, Université Paris 8 Vincennes-Saint Denis, France
Ninfa Criado Martínez, Universidad de Sevilla, España
María Dolores Gordón Peral, Universidad de Sevilla, España
Isabel María Íñigo Mora, Universidad de Sevilla, España
Manuel Maldonado Alemán, Universidad de Sevilla, España
Daniela Marcheschi, Università degli Studi di Perugia, Italia
Pedro Martín Butragueño, Colegio de México, México
Miguel Ángel Quesada Pacheco, Universitetet i Bergen, Norge
Angelica Valentinetti, Universidad de Sevilla, España
Alf Monjour, Universität Duisburg-Essen, Deutschland
María José Osuna Cabezas, Universidad de Sevilla, España
Fátima Roldán Castro, Universidad de Sevilla, España
Antonio Romano, Università degli Studi di Torino, Italia
Juan Pedro Sánchez Méndez, Université de Neuchâtel, Suisse
María Luisa Siguán Boehmer, Universitat de Barcelona, España
José Solís de los Santos, Universidad de Sevilla, España
Modesta Suárez, Université de Toulouse-Le Mirail, France
María Ángeles Toda Iglesia, Universidad de Sevilla, España
José Agustín Vidal Domínguez, Universidad de Sevilla, España
María Jesús Viguera Molins, Universidad Complutense de Madrid, España
Adamantía Zerva, Universidad de Sevilla, España

COMITÉ CIENTÍFICO

Juan Francisco Alcina Rovira, Universitat Rovira i Virgili, España
Gerd Antos, Martin-Luther-Universität Halle-Wittenberg, Deutschland
Gianluigi Beccaria, Università degli Studi di Torino, Italia
Isabel Carrera Suárez, Universidad de Oviedo, España
Carmen Herrero, Manchester Metropolitan University, England
Anna Housková, Univerzita Karlova, Česká Republika
Dieter Kremer, Universität Trier, Deutschland
Xavier Luffin, Vrije Universiteit Brussel, Belgique
Roberto Nicolai, Sapienza - Università di Roma, Italia
Marie-Linda Ortega, Université Sorbonne Nouvelle - Paris 3, France
Deborah C. Payne, American University, USA
Carmen Silva-Corvalán, University of Southern California, USA
Alicia Yllera Fernández, UNED, España

CONSEJO ASESOR

ESTUDIOS ÁRABES E ISLÁMICOS

Eva Lapiedra Gutiérrez, Universidad de Alicante, España

Pablo Beneito Arias, Universidad de Murcia, España

Carmelo Pérez Beltrán, Universidad de Granada, España

FILOLOGÍA ALEMANA

Georg Pichler, Universidad de Alcalá, España

Marta Fernández-Villanueva Jané, Universitat de Barcelona, España

María José Domínguez, Universidade de Santiago de Compostela, España

FILOLOGÍA CLÁSICA - LATÍN

Jesús Luque Moreno, Universidad de Granada, España

José Luis Moralejo Álvarez, Universidad de Alcalá de Henares, España

Eustaquio Sánchez Salor, Universidad de Extremadura, España

FILOLOGÍA CLÁSICA - GRIEGO

Didier Marcotte, Université Sorbonne Paris, France

Maurizio Sonnino, Sapienza-Università di Roma, Italia

Stefan Schorn, Université Catholique de Louvain, Belgique

FILOLOGÍA FRANCESA

Dolores Bermúdez Medina, Universidad de Cádiz, España

Monserrat Serrano Mañes, Universidad de Granada, España

María Luisa Donaire Fernández, Universidad de Oviedo, España

FILOLOGÍA ITALIANA

Giovanni Albertocchi, Universitat de Girona, España

Cesáreo Calvo Rigual, Universitat de València - IULMA, España

Margarita Borreguero Zuloaga, Universidad Complutense de Madrid, España

LENGUA ESPAÑOLA

Emilio Montero Cartelle, Universidade de Santiago de Compostela, España

Antonio Salvador Plans, Universidad de Extremadura, España

Antonio Briz Gómez, Universitat de València, España

LENGUA INGLESA

Emilia Alonso Sameño, Ohio University, USA

Carmen Gregori Signes, Universitat de València, España

Nuria Yanez-Bouza, Universidade de Vigo, España

LINGÜÍSTICA

Ángel López García, Universitat de València, España

Eugenio Martínez Celdrán, Universitat de Barcelona, España

Juan Carlos Moreno Cabrera, Universidad Autónoma de Madrid, España

LITERATURA ESPAÑOLA

Pedro M. Cátedra, Universidad de Salamanca, España

Flavia Gherardi, Università degli Studio di Napoli Federico II, Italia

Leonardo Romero Tobar, Universidad de Zaragoza, España

LITERATURA HISPANOAMERICANA

Teodosio Fernandez, Universidad Autónoma de Madrid, España

Noé Jitrik, Universidad de Buenos Aires, Argentina

Edwin Williamson, Oxford University, Inglaterra

LITERATURA INGLESA

Luis Alberto Lázaro Lafuente, Universidad Alcalá de Henares, España

Ricardo Mairal Usón, UNED, España

Carme Manuel Cuenca, Universitat de València, España

TEORÍA DE LA LITERATURA

José Domínguez Caparrós, UNED, España

Antonio Garrido Domínguez, Universidad Complutense de Madrid, España

Isabel Paraiso Almansa, Universidad de Valladolid, España

REVISORES DEL VOLUMEN 38, NÚMERO 2 (2024). ESTUDIOS LITERARIOS

Han actuado como revisores anónimos para uno o más artículos de este número, tanto los aceptados como los rechazados, los siguientes investigadores:

Albaladejo Mayordomo, Tomás (Universidad Autónoma de Madrid)
Albesa Pedrola, Elena (Universidad de Zaragoza)
Álvarez-Cifuentes, Pedro (Universidad de Oviedo)
Alvite Díez, María Luisa (Universidad de León)
Aradra Sánchez, Rosa María (Universidad Nacional de Educación a Distancia, UNED)
Aranda García, Nuria (Universidad de Zaragoza)
Barrera Velasco, Patricia (Universidad Internacional de La Rioja, UNIR)
Beneito Arias, Pablo (Universidad de Murcia)
Fernández Travieso, Carlota Cristina (Universidade da Coruña)
García Aguilar, Ignacio (Universidad de Córdoba)
García Sánchez-Migallón, Patricia (Universidad Complutense de Madrid)
Garrido Martín, Blanca (Universidad de Sevilla)
Isasi, Jennifer (Pennsylvania State University)
Malpartida Tirado, Rafael (Universidad de Málaga)
Monferrer Sala, Juan Pedro (Universidad de Córdoba)
Morillas García, Esther (Universidad de Málaga)
Moya García, Cristina (Universidad de Sevilla)
Muguruza Roca, María Isabel (Universidad del País Vasco)
Nieto Caballero, Guadalupe (Universidad de Extremadura)
Núñez Rivera, José Valentín (Universidad de Huelva)
Peña Martín, Salvador (Universidad de Málaga)
Pichel, Ricardo (Universidad Nacional de Educación a Distancia, UNED)
Toda Iglesia, María Ángeles (Universidad de Sevilla)
Tomasi, Giulia (Università di Trento)
Zaragoza Gómez, Verónica (Universitat de València)

ÍNDICE

| | |
|---|----|
| Sección Monográfica. Humanidades Digitales y Literatura Hispánica / <i>Digital Humanities and Spanish Literature</i> | 13 |
| Presentación / <i>Presentation</i> | 15 |
| Laura Hernández-Lorenzo (Universidad de Sevilla) | |
| Incidencia de las tecnologías digitales en la investigación en Literatura española del Siglo de Oro / <i>Impact of Digital Technologies on Research in</i> <i>Spanish Literature Of The Golden Age</i> | 21 |
| Sagrario López Poza (Universidade da Coruña) | |
| https://dx.doi.org/10.12795/PH.2024.v38.i02.01 | |
| Acerca de la sostenibilidad, archivo y preservación de proyectos de edición digital de textos hispánicos / <i>On the Sustainability, Archiving, and</i> <i>Preservation of Hispanic Digital Edition Projects</i> | 41 |
| Gimena del Rio Riande (Instituto de Investigaciones Bibliográficas y Crítica Textual, Consejo Nacional de Investigaciones Científicas y Técnicas) | |
| https://dx.doi.org/10.12795/PH.2024.v38.i02.02 | |
| Revolucionar el acceso al patrimonio librario: los sistemas de HTR entre humanidades digitales y ciencia de la información / <i>Revolutionizing</i> <i>Access to Library Heritage: HTR Systems between Digital Humanities</i> <i>and Information Science</i> | 59 |
| Stefano Bazzaco (Università di Verona) | |
| https://dx.doi.org/10.12795/PH.2024.v38.i02.03 | |
| El Proyecto Silem: Humanidades Digitales en las letras áureas / <i>The Silem</i> <i>Project: Digital Humanities in the Golden Letters</i> | 79 |
| Elena Cano Turrión (Universidad de Jaén) | |
| https://dx.doi.org/10.12795/PH.2024.v38.i02.04 | |

Los autores de las *Siete Partidas*: una visión desde la estilometría /
The Authors of the Siete Partidas: a Stylometric Approach 101
José Manuel Fradejas Rueda (Universidad de Valladolid)
<https://dx.doi.org/10.12795/PH.2024.v38.i02.05>

Autoría e identidad de género. Perspectivas cruzadas a propósito de
El pensionado de Santa Casilda / Authorship and Gender Identity. Crossed
Perspectives Regarding El Pensionado De Santa Casilda 131
Dolores Romero López (Universidad Complutense de Madrid)
María Jesús Fraga (Universidad Complutense de Madrid)
José Calvo Tello (Göttingen State and University Library)
<https://dx.doi.org/10.12795/PH.2024.v38.i02.06>

Sección Varia

Hipótesis sobre la biografía de Antonio de Torquemada y su obra literaria
a partir de nuevos documentos en los archivos históricos / *Hypothesis*
on the Biography of Antonio de Torquemada and his Literary Work Based
on New Documents in the Historical Archives 167
Jesús Fernando Cáceda Teresa (I.E.S. Valle del Cidacos - Calahorra,
La Rioja)
<https://dx.doi.org/10.12795/PH.2024.v38.i02.07>

La narrativa de Manuel Fernández y González frente al canon literario /
The Narrative by Manuel Fernández y González Facing the Literary Canon 183
Javier Muñoz de Morales Galiana (Universiteit Gent)
<https://dx.doi.org/10.12795/PH.2024.v38.i02.08>

La denuncia de estereotipos raciales y de género en la literatura juvenil:
las traducciones al español de *On the Come Up* (2019) y *Concrete*
Rose (2021) de Angie Thomas / *Denouncing Racial and Gender Stereotypes*
in Young Adult Literature: Translations into Spanish of on the Come Up
(2019) and Concrete Rose (2021) by Angie Thomas 199
Beatriz Rodríguez Rodríguez (Universidade de Vigo)
<https://dx.doi.org/10.12795/PH.2024.v38.i02.09>

Una nueva traducción castellana de las *Metamorfosis* de Ovidio (siglo xv) /
A New Castilian Traduction of the Ovid's Metamorphoses (15th Century) 219
Ana María Romera Manzanares (Universidad Complutense de Madrid)
<https://dx.doi.org/10.12795/PH.2024.v38.i02.10>

Reseñas de libros

- José Calvo Tello: *The Novel in the Spanish Silver Age. A Digital Analysis of Genre Using Machine Learning*. Beilefeld: Beilefeld University Press, 2021, 470 pp. ISBN: 978-3-8376-5925-2 249
Isabel Moyano Moreno (Universidad de Cádiz)
- Francisco Javier Blasco Pascual y Cristina Ruiz Urbón: *Análisis de textos desde la estilometría*. Salamanca: Editorial de la Universidad de Salamanca, 2022, 214 pp. ISBN: 978-84-1311-763-8 255
Patricia García Sánchez-Migallón (Universidad Nacional de Educación a Distancia)
- Ernesto Sierra: *Leopoldo Marechal y José Lezama Lima: Luces y sombras de la ciudad letrada*. Editorial Verbum, S. L., 2023, 196 pp. ISBN: 978-84-1337-955-5 259
Antonio Correa Iglesias (Universidad de Miami)
- José Teruel y Santiago López-Ríos, Eds.: *El valor de las cartas en el tiempo. Sobre epistolarios inéditos en la cultura española desde 1936*. Madrid: Iberoamericana Vervuert, 2023, 388 pp. ISBN: 9788491923664 263
Ellen Mayock (Washington & Lee University)
- Martín Rodríguez-Gaona: *Contra los influencers. Corporativización tecnológica y modernización fallida (o sobre el futuro de la ciudad letrada)*. Valencia: Pre-Textos, 2023, 474 pp. ISBN: 978-84-19633-26-2 267
Aitana Monzón Blasco (Universidad de Zaragoza)

Sección Monográfica
Humanidades Digitales y Literatura Hispánica
Digital Humanities and Spanish Literature

Laura Hernández-Lorenzo
Universidad de Sevilla
(Coord.)



ESTUDIOS LITERARIOS

REVOLUCIONAR EL ACCESO AL PATRIMONIO LIBRARIO: LOS SISTEMAS DE
HTR ENTRE HUMANIDADES DIGITALES Y CIENCIA DE LA INFORMACIÓN
REVOLUTIONIZING ACCESS TO LIBRARY HERITAGE: HTR SYSTEMS BETWEEN DIGITAL
HUMANITIES AND INFORMATION SCIENCE

STEFANO BAZZACO

Università di Verona

stefano.bazzaco.1@gmail.com

ORCID: 0000-0002-3496-3687

Recibido: 07-01-2024

Aceptado: 16-01-2024

RESUMEN

El presente trabajo busca ofrecer un estado de la cuestión sobre los recientes desarrollos en el campo de la transcripción automática de impresos antiguos y manuscritos con sistemas de HTR (Handwritten Text Recognition), fijando la atención prioritariamente en la creación reciente de modelos de HTR mixtos. Al respecto se explican las características principales de las herramientas más difundidas y el flujo de trabajo para la generación de modelos de reconocimiento de texto. En segundo lugar, se proporciona una muestra significativa de los modelos disponibles en la actualidad, insistiendo en el proceso de producción, los criterios adoptados y la evaluación de los resultados en relación con la experiencia madurada por el grupo de investigación Progetto Mambrino de la Universidad de Verona. Finalmente se proporcionan unas futuras pistas de investigación para la creación y difusión de estos recursos, haciendo hincapié en la necesidad de buscar una mayor sinergia entre contexto académico, expertos informáticos e instituciones de la memoria.

Palabras clave: Handwritten Text Recognition (HTR), modelos mixtos, Progetto Mambrino, ciencia de la información, edición digital académica.

ABSTRACT

The present work aims to offer a state of the art on recent developments in the field of automatic transcription of historical printed documents and manuscripts with HTR (Handwritten Text Recognition) systems, focusing primarily on the recent creation of HTR general models. In this regard, the main characteristics of the most widespread tools and the workflow for generating text recognition models are explained. Secondly, a significant sample of the models currently available is provided, insisting on the production process, the criteria

adopted and the evaluation of the results, in relation to the experience matured by the Progetto Mambrino research group of the University of Verona. Finally, some future research directions are provided for the creation and dissemination of these resources, emphasizing the need to seek greater synergy between the academic context, computer experts and memory institutions.

Keywords: Handwritten Text Recognition (HTR), general models, Progetto Mambrino, information science, digital scholarly edition.

1. INTRODUCCIÓN

La transcripción automática de documentos impresos y manuscritos ha representado desde siempre un sueño para los humanistas y los profesionales de la ciencia de la información por constituir un hito fundamental en la gestión de todo el material bibliográfico que ha sido volcado a la red de forma masiva durante las últimas décadas. Sin embargo, la empresa parecía de difícil alcance hace unos lustros, puesto que las tecnologías de reconocimiento de textos no aseguraban resultados fiables, sobre todo con impresos antiguos y manuscritos, y por consiguiente estas herramientas resultaban incapaces de tratar el problema frente al *big data* derivado de la digitalización en formato imagen de billones de obras de nuestro patrimonio escrito.

Más recientemente, los logros obtenidos en el campo de la inteligencia artificial sugieren que estamos más cercanos a este sueño. La idea de convertir una masa interminable de artefactos digitalizados en texto electrónico, fácil de buscar y procesable con dispositivos computacionales ha estado en el centro de las investigaciones más recientes en el campo de la informática, prometiendo alcanzar una nueva etapa de la textualidad digital¹.

El proceso del que hablamos presentaba unos problemas iniciales con que los especialistas debían necesariamente enfrentarse para que la empresa no fracasara o generara contenidos inservibles para la indexación y la búsqueda de carácter científico. Por una parte, las imágenes derivadas de la digitalización en larga escala realizadas en distintas etapas tecnológicas² no siempre alcanzaban un nivel de calidad suficiente para la aplicación de herramientas de transcripción automática. Por otra, los sistemas de transcripción automática, comúnmente llamados de Optical Character Recognition (de aquí en adelante OCR), no prometían resultados asequibles para que se pudiera imaginar una aplicación generalizada con los textos de nuestro patrimonio cultural digitalizado, con la consecuencia de que se generara cierta suspicacia con respecto a su empleo en distintos ámbitos, de las ciencias del documento al contexto académico (Cordell y Smith, 2018: 10-11; Bazzaco *et al.*, 2022: 75-76).

¹ Para ello, remito a la bibliografía presente en Bazzaco *et al.* (2022).

² Para más detalles, confróntese sobre este tema con el trabajo de Terras (2010), que se resume en Bazzaco *et al.* (2022).

Solamente en los últimos años, gracias al vertiginoso desarrollo de sistemas de Handwritten Text Recognition (de aquí en adelante HTR), que han reemplazado en su aplicación a las herramientas de OCR en las tareas de transcripción automática, se ha generado la sensación de que este proceso esté más al alcance. Por medio de *softwares* que consienten transcribir también imágenes en baja resolución sacadas de Google Books con un margen de error inferior al 5%, lo cual era impensable en los albores del nuevo milenio, las dificultades iniciales que comentamos han ido desalentándose, y con ellas también la reticencia por parte de muchos especialistas. No obstante, el avance en el uso de estas herramientas está generando unas complejidades inéditas a la hora de aplicarse concretamente al enorme conjunto de textos digitalizados, complejidades que los especialistas deben de tomar en consideración para que no se perjudique o alargue el entero proceso de conversión, que debe nutrirse de colaboraciones y buenas prácticas comunes entre los distintos grupos que se ocupan de la difusión, gestión e investigación del patrimonio bibliográfico.

A partir de estas premisas, el presente trabajo tiene el objetivo de sustentar cierta sinergia entre los profesionales de la información y el contexto académico. En concreto, se valora el estado del arte de los sistemas de HTR y se describe el flujo de trabajo que generalmente está en la base de las herramientas más difundidas. En segundo lugar, se subraya el interés que en este campo están obteniendo los modelos de HTR mixtos y se proponen soluciones viables para su elaboración, sustentando nuestro discurso en la experiencia científica y formativa madurada por el grupo de investigación Progetto Mambrino de la Universidad de Verona en los marcos del proyecto de relevante interés nacional (PRIN) «Mapping Chivalry: Spanish Romances of Chivalry from Renaissance to 21st Century. A digital approach» (2018-2023)³ y del Progetto di Eccellenza «Le Digital Humanities applicate alle Lingue e Letterature Straniere» (2018-2022)⁴. Finalmente, se describen las propuestas experimentadas en el ámbito de las humanidades digitales y se vislumbran las posibles implicaciones para filólogos, técnicos informáticos y especialistas de la información, sentando las bases para futuras colaboraciones entre contextos que todavía sufren cierta falta de permeabilidad (Alvite-Díez y Barrionuevo, 2020).

³ El proyecto incluye cuatro unidades distintas que han desarrollado sus propias bases de datos: la de Verona (Biblioteca Digital del Progetto Mambrino, dir. A. Bognolo y S. Neri: mambrino.mappingchivalry.dlcs.univr.it), la de Trento (MemoRam, dir. C. Demattè y G. Tomasi: <http://memoram.mappingchivalry.dlcs.univr.it>), la de Salerno (Teatro Caballeresco, dir. D. Crivellari: <http://teatrocaballeresco.mappingchivalry.dlcs.univr.it>) y la de Roma La Sapienza (AmadisSigloXX, dir. E. Sarmati: <http://amadissigloxx.mappingchivalry.dlcs.univr.it>) (cons. 17/01/2024).

⁴ <https://www.univrmagazine.it/2023/07/18/lingue-e-letterature-straniere-dipartimento-di-eccellenza/> (cons. 08/07/2024)

2. SISTEMAS DE OCR/HTR Y FLUJO DE TRABAJO

La historia de los sistemas de reconocimiento de texto es muy larga y puede retrotraerse hasta las iniciales experimentaciones en el campo del *automatic recognition* (Bazzaco *et al.*, 2022: 73); sin embargo, es a partir de los años 90 cuando se empieza a emplear *software* de OCR para la transcripción de los artefactos derivados de la digitalización masiva de documentos. Estas herramientas, como sugiere el nombre, permiten convertir las imágenes del documento escrito en texto electrónico, es decir, *machine readable*, medible y procesable por parte de la máquina, a partir de la separación y detección de cada uno de los caracteres contenidos en el documento procesado (Terras, 2022a: 181; Terras, 2022b: 138).

Las bibliotecas en línea como archive.org y los proyectos de periódicos digitales adoptaron estos sistemas porque permitían la búsqueda de metadatos en repositorios textuales muy vastos, pero fue la escasa fiabilidad de los resultados con textos impresos antiguos y manuscritos la que limitó su aplicación a larga escala. El germen del problema se identificaba con una característica primordial del OCR, es decir, la de tratar cada carácter de la fuente de forma aislada en una relación 1:1 con su correspondiente carácter transcrito. De hecho, en los impresos antiguos y en los manuscritos con frecuencia los caracteres no aparecen aislados, sino unidos por ligaduras, y su representación es variable también dentro de un mismo documento; sin contar que con frecuencia aparecen signos gráficos particulares (como el signo tironiano y las abreviaturas) que no proponen una relación unívoca entre cada realización gráfica y la correspondiente transcripción.

Si bien existen sistemas de OCR como OCRopus (<https://github.com/ocropus>) y Tesseract (<https://github.com/tesseract-ocr/tesseract>) que, aprovechando los recientes logros en el campo del *machine learning* han permitido transcribir algunos textos manuscritos, son raros los casos de aplicación exitosos porque por su misma naturaleza estas herramientas prefieren tratar caracteres separados por espacios blancos y uniformes. De ello, Terras deriva que «los sistemas de OCR todavía tienen problemas con grafías, diseños o soportes complejos» (Terras, 2022a: 182, traducción mía), como ha sido demostrado por Mancinelli (2016) y Cordell y Smith (2018).

Solamente gracias al desarrollo de las tecnologías de entrenamiento profundo (*deep learning*), los sistemas de HTR han llegado a ser las herramientas preeminentes en el campo de la transcripción automática. Estos *softwares* no se fijan en caracteres aislados, sino en las palabras y líneas que componen un texto: a partir de ellas extraen las características visuales y buscan calcular las probabilidades de solapamiento con realizaciones gráficas de otras porciones de texto. Para su desarrollo ha sido significativa la implementación de redes neuronales recurrentes (RNN), clases de algoritmos que funcionan como las redes cerebrales de los seres humanos y que toman decisiones pasando por distintas capas de procesamiento o estratos hasta analizar el contenido de una imagen y buscar recurrencias en la disposición de los

elementos que la componen⁵. Se trata, en otras palabras, de tecnologías que aseguran que el ordenador, a partir de un entrenamiento preliminar, pueda clasificar en *clusters* las distintas realizaciones gráficas de porciones textuales y predecir qué caracteres les corresponden en el texto transcrito.

Existen hoy en día distintas soluciones de transcripción automática que aprovechan la inteligencia artificial para el reconocimiento de texto, pero se diferencian entre ellas por el enfoque adoptado. En ocasiones, las experimentaciones pueden darse dentro de proyectos con necesidades específicas, y por lo tanto nacen de la colaboración entre expertos informáticos y especialistas de la información, proveyendo soluciones *ad hoc* y situándose en un horizonte limitado de reutilización. En otros casos se prefiere apoyarse en iniciativas comerciales para la gestión del flujo de digitalización, por ejemplo, las grandes editoriales y sus plataformas tecnológicas, que evidentemente no fomentan la cooperación y la reproducibilidad de los resultados obtenidos. Finalmente, existe un tercer planteamiento que se apoya en los productos generados por la misma comunidad científica, es decir, *softwares* que nacen en el contexto de proyectos de largo alcance y que se convierten en proveedores de servicios para todo tipo de usuario interesado. El resultado es que, a pesar de que el campo del reconocimiento de texto haya experimentado notables avances en los últimos años, falta todavía una estandarización en los distintos planteamientos que gobiernan su desarrollo.

Las colaboraciones entre instituciones de la memoria, investigadores y especialistas computacionales han ofrecido resultados interesantes: es el caso de proyectos como *In codice ratio*, para el análisis y el procesamiento de las informaciones contenidas en los Archivos Secretos del Vaticano (Firmani *et al.*, 2018), o el sistema Monk, desarrollado por la Universidad de Groningen, el Instituto de Inteligencia Artificial ALICE y el Dutch National Archive, que proporciona recursos de anotación, reconocimiento y búsqueda textual basados en algoritmos que van más allá de las posibilidades ofrecidas por los tradicionales sistemas de OCR (Weber *et al.*, 2018)⁶.

Al contrario, las grandes editoriales y las plataformas tecnológicas de largo alcance se sirven de herramientas de HTR para la pesquisa de enteras colecciones digitalizadas, como por ejemplo hacen Adam Matthew Digital y Gale (Terras, 2022a: 184), o proveyendo *softwares* bajo licencia, como es el caso de *Fabricius*, la reciente

⁵ La web de IBM ofrece informaciones interesantes al respecto: <https://www.ibm.com/it-it/topics/recurrent-neural-networks> (cons. 27/12/2022). Sobre su aplicación en el campo de la transcripción automática, considérense Mühlberger *et al.* (2019: 968-969), Schwarz-Ricci (2022: 172-173) y Bermúdez Carreño (2023: 111-112).

⁶ La propuesta es de interés porque permite la explotación en abierto de los algoritmos de reconocimiento generados a lo largo del desarrollo del proyecto; sin embargo, el enlace al *search engine* de Monk no parece en la actualidad estar disponible para el usuario. Consúltese la web del proyecto, disponible en el siguiente enlace: <https://www.ai.rug.nl/~lambert/Monk-collections-english.html> (cons. 27/12/2022).

tecnología de Google para el reconocimiento de jeroglíficos por medio de IA⁷. Sin embargo, estos sistemas comerciales «pueden ser opacos, los algoritmos y recursos empleados raramente se publican, y hay que cuidar cuestiones de *copyright*, licencia de las imágenes y almacenamiento a largo plazo de los recursos digitales cuando se establecen alianzas con estas grandes entidades tecnológicas» (Terras, 2022a: 184, traducción mía).

En el extremo opuesto se sitúan los sistemas de HTR elaborados y gestados en el contexto de la comunidad científica, es decir, *Virtual Research Environments* que hacen accesible esta tecnología a cualquier tipo de público, desde los estudiosos de humanidades a los bibliotecarios y archiveros. Se trata de herramientas de acceso abierto que facilitan la transcripción manual de forma colaborativa y después consienten emplear esa transcripción para el aprendizaje de la máquina, que será capaz de interpretar otras porciones textuales cercanas en su realización gráfica a la base producida por el usuario, permitiendo en una fase posterior ejecutar búsquedas y extracción de información. Como señalan Souibgui *et al.* (2022: 4), estas tecnologías son todavía escasas en número, pero contamos con dos infraestructuras principales: Transkribus (READ Coop, <https://readcoop.eu/it/transkribus/>) y eScriptorium (<https://gitlab.com/scripta/escriptorium>). Los dos *softwares* ofrecen unas posibilidades parecidas, ya que proporcionan una interfaz gráfica (GUI) de notable sencillez y permiten que el usuario ejecute el entrenamiento y reconocimiento con HTR sobre sus textos de interés.

Transkribus y eScriptorium permiten transcribir impresos antiguos y documentos manuscritos con un margen de error muy bajo por medio del aprendizaje ejecutado por el usuario. El flujo de trabajo en resumidas cuentas prevé: la importación de imágenes digitalizadas en la plataforma; la segmentación semi-automatizada de cada página en distintas zonas de texto, como regiones y líneas (*Layout Analysis* o *Layout Detection*); la transcripción manual, individual o colaborativa, de porciones de texto, que comúnmente se llama *Golden Standard Transcription* o *Ground Truth*. La transcripción manual constituye la base del entrenamiento del *software* de HTR para la creación de unos modelos de reconocimiento. Entonces, a partir de la producción de un modelo adecuado, las herramientas permiten transcribir la restante parte del documento de interés del usuario y exportar el resultado en formatos estándar, como TXT, DOC o XML⁸.

⁷ Más informaciones en la web de *Fabricius*, gestionada por Google Arts & Culture: <https://artsandculture.google.com/experiment/fabricius/gwHX41Sm0N7-Dw> (cons. 27/12/2023). En relación con los sistemas de HTR, citan la herramienta Souibgui *et al.* (2022: 4) y Terras (2022a: 184).

⁸ Para una descripción detallada del funcionamiento de Transkribus remito a Mühlberger *et al.* (2019), Bazzaco (2018, 2020) y Bazzaco *et al.* (2022). Sobre el flujo de trabajo de eScriptorium, consúltese Pinche (2023) y la documentación disponible al siguiente enlace: <https://lectaurep.hypotheses.org/documentation/prendre-en-main-escriptorium> (cons. 27/12/2023).

Lo que diferencia mayormente las dos plataformas es el planteamiento teórico a la base de cada una, puesto que eScriptorium se apoya en la infraestructura de Kraken (<https://kraken.re/main/index.html>) y sigue una filosofía completamente Open Access, mientras que Transkribus no proporciona el acceso a los modelos de reconocimiento y desde 2020 prevé la adquisición de créditos para ejecutar el reconocimiento con HTR. Se trata de un asunto controvertido que ha suscitado polémicas y disensiones por parte de los especialistas de las HHDD, sostenedores de infraestructuras totalmente abiertas⁹. No obstante, hay que reparar en el hecho de que en la actualidad Transkribus sigue contando con una comunidad más amplia de usuarios, lo cual tiene efectos inmediatos sobre los resultados obtenidos¹⁰, y es más eficiente, ya que necesita de una potencia de cálculo inferior por parte del usuario, apoyándose en unos servidores colocados en la Universidad de Innsbruck.

3. POTENCIALIDADES DE LA TRANSCRIPCIÓN AUTOMÁTICA: LOS MODELOS MIXTOS

Como comentamos en las páginas anteriores, los *softwares* de HTR consienten obtener resultados a veces muy fiables con impresos antiguos y textos manuscritos, y esto está acrecentando el interés por parte de los especialistas, sobre todo los de humanidades, que vislumbran un crecimiento exponencial en el desarrollo de sus proyectos digitales. No obstante, sigue siendo problemático el asunto de determinar la concreta eficiencia de los modelos entrenados sobre un texto individual en relación con documentos que no han constituido parte del entrenamiento, sobre todo en el caso de documentos manuscritos donde aparecen diversas grafías, a veces muy diferentes entre ellas (Hodel *et al.*, 2021: 3). Esto pasa porque los sistemas de HTR por su misma naturaleza aprenden de lo que está incluido en el *dataset* (o sea, parte de la transcripción manual *Golden Standard*), pero de entrada tienen más dificultades en prever la disposición gráfica del texto con documentos que para ellos siguen siendo desconocidos.

De hecho, cuando se hace referencia al margen de error porcentual de un modelo de HTR individual se está tomando en consideración un dato potencial, supuestamente difícil de determinar a la hora de aplicar el modelo a otro texto. Esto pasa porque la valoración de la eficiencia de cada modelo se basa en una interpretación

⁹ La aplicación del sistema de pago se comunicó a los usuarios por medio de un anuncio web, hoy indisponible (Terras, 2022a: 185, nota 36). Sin embargo, quizás por las críticas recibidas, desde enero de 2024 la Cooperativa Europea READ propone distintas tipologías de planes, que van del «individual», gratuito, pero con restricciones (límite de 5 entrenamientos por mes, baja prioridad), a planes del tipo «scholar» o «organization», que son de pago y aseguran funcionalidades avanzadas y un mayor soporte. Para más informaciones acerca de la gestión de planes y créditos, consúltese la siguiente página web: <https://www.transkribus.org/plans> (cons. 11/07/2024).

¹⁰ Considérese que, según las bases del *machine learning*, la máquina aprende del procesamiento de cualquier documento, cualquiera que sea su relación con los nuevos materiales que se les proporciona. De tal manera, la aportación de distintos investigadores en el globo, sin distinciones de tipología o idioma, incrementa las capacidades de la plataforma para toda la comunidad de usuarios.

a veces equívoca. Para el aprendizaje se le pide al usuario que solo una parte de su transcripción manual sea la base del entrenamiento (*Training Set*), mientras que otra porción, normalmente el 10%, vaya a constituir el conjunto textual de validación (*Validation Set*): el índice de fiabilidad del modelo (*Character Error Rate* o CER) se basa pues en la distancia entre el texto contenido en el conjunto de validación y el texto que el modelo reconocería al aplicarse a esos mismos materiales, es decir, cuántos cambios, añadiduras y supresiones se necesitarían para llegar de uno a otro (*edit distance*). Es por lo tanto lógico que el índice de error se fundamente solo en las páginas incluidas en el set de validación, y no tenga en cuenta —¡cómo podría!— la infinita gama de posibles aplicaciones a porciones del mismo texto o, exagerando, a documentos distintos. El CER es por esta razón un dato eventual, cuando no accidental, porque depende de un conjunto de páginas de validación, a veces muy exiguo, que se toma como caso ejemplar.

El resultado es que no se puede determinar con anterioridad el comportamiento del algoritmo en contextos distintos cuando se le somete a la máquina un nuevo documento por interpretar, y esto, en el caso de que se tenga la intención de aprovechar modelos de HTR públicos creados por otros investigadores, conlleva dificultades a la hora de elegir cuál es el modelo individual adecuado, que prometa resultados más cercanos a los propósitos del usuario. Para valorar de forma contundente la consistencia del modelo habría que testear su eficiencia en distintos casos de estudio (*samples*), por ejemplo, documentos de la misma época y redactados en una grafía parecida, que constituyan una muestra significativa para sus futuras aplicaciones.

Con el fin de sortear este límite intrínseco de evaluación de modelos individuales se sugirió inicialmente emplear el modelo como base para otros aprendizajes específicos (*fine tuning*). En tal caso, el modelo originario representaría la base (*base model*) de un modelo individual enriquecido con otros materiales de entrenamiento, para cuya creación se necesita volver a la transcripción manual. Sin embargo, esta tarea no solamente vuelve a ser engorrosa en términos de tiempo y trabajo, sino también difícilmente puede asegurarse que el proceso no tenga que repetirse cada vez que se necesite transcribir un texto distinto. Además, ha sido demostrado que los modelos basados en *datasets* muy extensos de textos del mismo tipo llevan a una especialización del modelo mismo, con una correspondiente degeneración de resultados con otros textos desconocidos (Hodel *et al.*, 2021: 5-6)¹¹.

La propuesta más interesante para generalizar la aplicación del HTR corresponde con la creación de modelos mixtos (en inglés: *general models*), es decir, oportunamente compuestos por un conjunto heterogéneo de documentos transcritos. Tales recursos pueden ser de distinta tipología, dependiendo de su composición y de los materiales de partida. En el caso de los impresos antiguos, la tendencia es de incluir

¹¹ El fenómeno se conoce como *overfitting*. Al respecto véase Rabus (2019: 11-12).

en un único modelo realizaciones distintas de una misma tipografía, por ejemplo, textos cercanos compuestos con un mismo tipo u obras salidas de un mismo taller en un lapso temporal determinado¹². En el caso de los textos manuscritos, por otra parte, se prefiere trabajar con textos de un mismo autor, de modo que el algoritmo se fundamente en distintas realizaciones de una grafía específica, que a pesar de que provengan de una misma mano difieren por varias razones en su aspecto, debido a las condiciones físicas de escritura y los diversos soportes. Asimismo, en ambas situaciones se consigue una ulterior variedad interna del modelo al incluir materiales digitalizados de distinta procedencia, puesto que varían las técnicas de conservación digital y las tecnologías de escaneo empleadas.

Por ser muy eficaces y adaptables a múltiples circunstancias, también en presencia de documentos muy breves que no proporcionarían material suficiente para el entrenamiento¹³, los modelos mixtos están en el centro de los intereses de la comunidad científica. Sin embargo, para que el proceso de creación de los mismos se emprenda de forma sistemática es necesario fundamentar cada experimentación en unas prácticas comunes, que apunten a la sostenibilidad y congruencia del trabajo preparatorio.

4. MODELOS DE HTR MIXTOS PARA IMPRESO ANTIGUOS EN ESPAÑOL. ESTADO DE LA CUESTIÓN A LA LUZ DE LA EXPERIENCIA DEL PROGETTO MAMBRINO (UNIVERSITÀ DI VERONA)

En el caso de la transcripción automática de impresos antiguos se puede afirmar que la tecnología HTR ha llegado a una fase madura, puesto que existen modelos de reconocimiento fiables que alcanzan un margen de error inferior al 3% también con textos que no forman parte del *dataset* de entrenamiento. En la plataforma Transkribus, por ejemplo, se encuentran algunos modelos mixtos al alcance de cualquier usuario registrado (*Public Models*), que ofrecen una buena muestra de las potencialidades del HTR.

El modelo «Spanish Golden Age Prints 1.0», creado en el marco del proyecto ETSO (Estilometría aplicada al Teatro del Siglo de Oro), está basado en impresos teatrales del Siglo de Oro y cuenta con un *dataset* de 74 129 palabras transcritas. Se trata de uno de los primeros intentos de reconocimiento de la tipografía redonda: la transcripción manual respeta la ortografía de los textos originales y el modelo alcanza un margen de error de 0,91% sobre el conjunto de validación, que es una tasa de acierto notable, sobre todo para la transcripción automática de documentos con una misma tipografía (Cuéllar, 2023: 107). Sin embargo, la intención de los

¹² Se siguió la primera perspectiva en la creación de modelos para la gótica y la redonda españolas del XVI, así como descrito en Bazzaco *et al.* (2022).

¹³ El problema es tratado por Hodel *et al.* (2021: 3). Con referencia al contexto ibérico, este mismo asunto interesa por ejemplo la transcripción automática de relaciones de sucesos, como se sugiere en Bazzaco *et al.* (2022).

investigadores principales del proyecto es la de testear el *software* para generar transcripciones que se puedan comparar entre ellas por medio de herramientas computacionales de estilometría, para luego buscar correspondencias entre textos cercanos y aventurarse en atribuciones de autoría. Para ello, textos que presentan formas distintas de una misma palabra no son útiles porque llevarían a interpretaciones erróneas, silenciando o enfatizando la cercanía entre las obras del corpus estudiado (2023: 107-108). Por lo tanto, los responsables han entrenado también otro modelo con transcripciones modernizadas, sirviéndose de la función de alineación automática del texto contenido en las imágenes con obras ya editadas. Este segundo modelo de HTR llamado «Spanish Golden Age Prints (Spelling Modernization) 1.0» está compuesto por 2 757 908 palabras y llega a un CER del 3,10%, pero lógicamente es menos fiable del precedente por la complejidad de la tarea que se le pide a la máquina, es decir, la de interpretar contextualmente los signos gráficos presentes en la página y tomar decisiones de forma no unívoca, añadiendo, suprimiendo y alterando los caracteres visibles¹⁴. A pesar de estas limitaciones, el modelo se empleó en ETSO para generar transcripciones de impresos de forma no supervisada, proporcionando unos resultados relevantes en el campo de la atribución de textos teatrales del XVII-XVIII¹⁵.

Por otra parte, el uso de sistemas de HTR se extiende también a otros campos de la gestión de la información: en el marco del proyecto CLARA-HD (<https://clara-nlp.uned.es/home/dh/>) de la UNED, por ejemplo, se empleó Transkribus en relación con técnicas de procesamiento del lenguaje natural (PNL). Al respecto, debemos a Eva Sánchez-Salido y Ana García-Serrano la creación de otro modelo público llamado «Diario de Madrid 1788-1825», esta vez basado en la interpretación de periódicos históricos de los siglos XVIII, XIX y principios del XX (Menta *et al.*, 2022). El entrenamiento en este caso fue llevado a cabo sobre 91 640 palabras, pero las investigadoras emplearon también como *base model* el primer modelo HTR de Cuéllar, perfeccionado hasta alcanzar la tasa de error del 1% sobre el conjunto de validación para los documentos objeto de estudio del proyecto. A partir de los resultados de transcripción obtenidos, el objetivo en este caso es la detección de entidades como localizaciones y profesiones mencionadas en los artículos, según los temas de interés de los historiadores implicados (Menta *et al.*, 2022: 73).

En el marco del Progetto Mambrino de la Universidad de Verona se empezó la experimentación con Transkribus para la transcripción automática de impresos italianos en cursiva (Bazzaco, 2018), pero muy pronto se decidió también emplear el *software* para el reconocimiento de documentos españoles en gótica y redonda (Bazzaco, 2020; Bazzaco *et al.*, 2022). Por medio de la formación de un grupo de trabajo

¹⁴ Las experimentaciones llevadas a cabo por Cuéllar son de sumo interés también porque nos sitúan en el centro de un debate sobre hasta qué punto es posible modernizar las transcripciones durante la fase de entrenamiento del modelo. Al respecto véase también Bazzaco *et al.* (2022: 94).

¹⁵ Por ejemplo, consúltense García-Reidy (2019) y Kroll y Sanz-Lázaro (2022).

que reunía investigadores de distinta procedencia se pudieron crear dos distintos modelos de HTR, «SpanishGothic_XV-XVI_extended» (220 904 palabras de entrenamiento; CER set de validación 0,91%) y «SpanishRedonda_sXVI-XVII_extended» (119 856 palabras de entrenamiento; CER set de validación 1,30%), actualmente de dominio público en la plataforma¹⁶.

Las necesidades del proyecto colaborativo que lanzamos en ese momento eran de naturaleza distinta de los dos casos citados anteriormente porque el propósito principal era el de poner a prueba la herramienta para generar textos que pudiesen ser explotados para la creación de ediciones digitales académicas¹⁷. De hecho, Transkribus es un ambiente de trabajo colaborativo eficaz, con funcionalidades muy útiles a la hora de emprender un trabajo de este tipo, por ejemplo, porque posibilita la extracción de ficheros XML-TEI en que se preservan las coordenadas de líneas del documento fuente y las etiquetas semánticas asociadas al texto transcrito. Sin embargo, nuestra intención era la de proporcionar un texto que seguía criterios de transcripción rigurosos, fácilmente explotable por parte del filólogo. Decidimos por lo tanto crear dos modelos mixtos muy especializados partiendo de transcripciones manuales semi-diplomáticas que respetaran en lo posible los elementos tipográficos presentes en la página, con la única excepción de la «s larga» (f) y las abreviaturas, que decidimos desarrollar con el fin de facilitarle un texto limpio al usuario final.

Los dos modelos de HTR obtenidos son bastante consistentes y aseguran obtener transcripciones con tasas de error inferiores al 1% con la mayoría de los textos impresos que presentan tipos parecidos a los que se incluyeron en el *dataset*, compuesto por obras muy variadas en términos de género, taller de procedencia y técnicas de adquisición digital¹⁸. Para evitar el fenómeno de *overfitting* decidimos seleccionar una porción textual determinada de cada obra (alrededor de 20 páginas), de manera que ninguna de ellas prevaleciera sobre las otras. Además, se fijaron las pautas para una constante puesta al día de los modelos, con el reclutamiento de jóvenes investigadores en los futuros talleres y su sucesiva inclusión como colaboradores del proyecto.

Los resultados obtenidos inspiraron a otros investigadores para el desarrollo de sus propias iniciativas. Al respecto, el caso más significativo es el del proyecto 7PartidasDigital (<https://7partidas.hypotheses.org/>) de la Universidad de Valladolid, cuyo objetivo es la edición crítica digital de la obra legislativa de Alfonso X el Sabio

¹⁶ Para más detalle sobre el proyecto colaborativo y la constitución de los *datasets*, consúltese Bazzaco *et al.* (2022: 93-96).

¹⁷ Con respecto al campo de la edición digital académica la bibliografía es extensiva: de entrada, para el contexto español, señalamos las publicaciones de Allés Torrent (2020) y Alvite-Díez y Rojas Castro (2022).

¹⁸ Se detalla la composición de los dos modelos en Bazzaco *et al.* (2022: 99-121). Consúltese también al respecto la descripción presente en la web de Transkribus en la sección «Public Models»: <https://recoop.eu/transkribus/public-models> (cons. 27/12/2023).

a partir de la transcripción automática de todos los testimonios y su anotación en XML-TEI. En relación con estas tareas, el profesor Fradejas Rueda está produciendo distintos modelos de reconocimiento para la letra impresa¹⁹, adaptando nuestros criterios de transcripción a los del Hispanic Seminar of Medieval Studies (HSMS) e integrando herramientas de postcorrección basadas en diccionarios controlados, para comparar finalmente los testimonios transcritos con tecnologías de colación automática (Fradejas Rueda, 2022).

La gran variedad de los proyectos señalados demuestra cómo el campo de la transcripción automática de impresos antiguos con sistemas de HTR es un campo extremadamente productivo, y, sin embargo, todavía en evolución y susceptible de mejora en relación con las distintas tareas del filólogo. No obstante, con el fin de valorar el trabajo hecho, señalamos un punto fundamental que comparten todos los proyectos mencionados, es decir, la intención de proveer una descripción detallada de los modelos de HTR producidos que se basan en criterios de transcripción certificados y en corpus bien definidos, lo cual puede favorecer en el futuro su empleo por parte de otros especialistas.

5. EXPERIMENTACIONES CON LETRA MANUSCRITA EN ESPAÑOL

Con respecto al reconocimiento de textos impresos, la transcripción automática de documentos manuscritos sigue siendo un ámbito todavía en desarrollo que necesita apoyarse en ulteriores fases de experimentación y evaluación de las posibilidades ofrecidas por las herramientas disponibles. En efecto, los problemas que surgen a la hora de aplicar sistemas de HTR a grafías manuscritas de difícil interpretación, que se presentan muy variables también dentro de un mismo documento, ya que resienten de las condiciones de escritura y de la calidad y tipología de los soportes, están todavía lejos de solucionarse.

Actualmente existe un número muy parco de modelos de HTR para manuscritos españoles al alcance de los investigadores.

Por lo que atiene a la plataforma eScriptorium, se ha llevado a cabo una sola experimentación con textos peninsulares a cargo de Gille Levenson (2023: 2), quien se ocupó de la creación de un modelo a partir de la transcripción de unos 300 folios pertenecientes a diez manuscritos y un incunable del siglo xv del *Regimiento de Príncipes*. En concreto, el autor documenta de forma admirable el trabajo llevado a cabo, indicando que se empleó directamente el *software* Kraken (versión 4.2.0, <https://github.com/mittagessen/kraken/releases/tag/4.2.0>) para operaciones de refinamiento del modelo. De hecho, se asegura que el procesamiento proporcionó unos resultados de transcripción automática muy buenos, inferiores al 10% de tasa

¹⁹ Para una muestra de los modelos de HTR que se están desarrollando en el marco del proyecto 7PartidasDigital, consúltense el siguiente enlace: https://7partidas.hypotheses.org/11531#footnote_6_11531 (cons. 27/12/2023).

de error, no solamente con las obras incluidas en el *dataset*, sino también con textos *out-of-domain*; sin embargo, la propuesta conlleva cierta competencia técnica que no siempre está al alcance de los estudiosos de humanidades y que quizás representa un límite a la hora de valorar concretamente la repercusión que puede tener la labor llevada a cabo.

Por otro lado, también Transkribus ha sido empleado para el reconocimiento de la letra manuscrita. En la plataforma existen solamente dos modelos públicos. El primero de ellos ha sido realizado por Cuéllar siempre en relación con el proyecto ETSO y presenta, a grandes rasgos, las mismas características del modelo para impresos con modernización gráfica, es decir, que está basado en una cantidad enorme de textos teatrales manuscritos (recoge 3 250 116 palabras), pero diversamente de este no alcanza resultados fiables, ya que al tratar textos manuscritos el CER sube notablemente, hasta alcanzar el 9,20%, lo cual implica que los resultados de los análisis estilométricos con textos desconocidos deben ser tomados con considerables precauciones (Cuéllar, 2023: 112). El segundo modelo público, disponible desde septiembre de 2021, está basado en la grafía del emperador Carlos V a la altura del año 1543 y fue realizado a partir de la edición del volumen de Ball y Parker (2014) sobre las instrucciones al joven príncipe Felipe II. El trabajo ha sido llevado a cabo según los criterios editoriales establecidos por los dos editores e incluye la transcripción de una carta y algunos textos políticos-institucionales, pero falta una descripción detallada del proceso de producción y tan solo se publicaron tres ejemplos de las imágenes digitalizadas empleadas por el proyecto²⁰, lo cual no consiente suponer posibles reutilizaciones del modelo. Por consiguiente, se puede afirmar que el modelo de HTR creado, por estar basado en un periodo cronológico limitado y presentar pocas variaciones internas, puede considerarse con respecto a su empleo muy cercano a un modelo individual, más útil probablemente para la implementación de modelos mixtos que para la explotación por parte de otros investigadores.

Si bien no aparecen otras indicaciones en la plataforma, sabemos que existen otros proyectos que usan Transkribus para la interpretación de textos manuscritos. Por ejemplo, en el contexto del proyecto 7PartidasDigital, que como recordamos trata de transcribir todos los ejemplares de la obra alfonsí según el sistema de transcripción del HSMS, se documenta el intento de reconocer la gótica librería del XIII a partir del entrenamiento con distintos manuscritos, empezando por el manuscrito VITR/4/6 de la Biblioteca Nacional de España, con un aprendizaje que llevó a una tasa de error del 3,90%²¹. Asimismo, en el interior de un proyecto

²⁰ Al respecto, véase la descripción del proyecto disponible en Zenodo: <https://zenodo.org/records/5381739> (cons. 27/12/2023).

²¹ Más informaciones están disponibles en el siguiente enlace: https://7partidas.hypotheses.org/11531#identif_1_11531 (cons. 27/12/2023). Se relata también que entre los objetivos del proyecto reside la idea de «crear un macromodelo para manuscritos medievales del iberorrománico central.

de arqueología dedicado a la transcripción del *Catálogo Monumental de España*, consta que se produjeron tres distintos modelos de HTR para el reconocimiento de la letra manuscrita de principios del xx, basándose en tres catálogos de Ávila, Soria y Burgos, respectivamente redactados por Manuel Gómez-Moreno, Juan Cabré y Narciso Sentenach (Liceras Garrido *et al.*, 2022: 273). También en este caso se ofrece un modelo mixto llamado «Early20thcentury» que integra las tres grafías en objeto y alcanza un CER del 5%, representando un primer paso en la creación de un modelo mixto para manuscritos de esa época (Liceras Garrido *et al.*, 2022: 274).

Además, en la actualidad se está registrando una nueva tendencia que prevé la creación de macromodelos muy extensos que abarcan textos impresos y manuscritos de épocas muy distantes. Es el caso de «Coloso Español», un modelo de HTR entrenado en la plataforma Transkribus y disponible desde agosto de 2023, que integra todos los modelos que presentamos en estas páginas y algunos más, sin distinción de tipología, periodo, (tipo)grafía y criterios de transcripción. La creación de este recurso ha sido coordinada por Álvaro Cuéllar con la participación de los investigadores que realizaron los modelos que contiene, que concedieron la libre incorporación de sus *datasets* para el aprendizaje²²: de momento, esta es la única información técnica que se suministra, junto con la sorprendente tasa de error de 3,90% para el conjunto de validación.

Sin embargo, las posibles aplicaciones y los resultados que se pueden obtener con estos macromodelos son un terreno todavía por explorar. Lo que se puede afirmar de momento es que se trata de propuestas que están en el centro de los intereses de la comunidad científica por las posibilidades que potencialmente ofrecen, pero que, a semejanza de los recientes *softwares* de Large Language Models (LLM), priman la cantidad sobre la calidad, lo cual no es un mal en sí, a condición de que se alerte el usuario sobre lo que puede sacar de ellos²³. En el caso de que se quiera ejecutar una

Es un modelo ambicioso porque la lengua medieval de ese ámbito peninsular incluye las variedades leonesa (con rasgos gallego-portugueses), castellana (en su amplio abanico), riojana, navarra y aragonesa (con influencia catalana); el espacio temporal abarca desde el reinado de Fernando III (1199-1252) hasta los Reyes Católicos (fines del siglo xv) y hay un amplio catálogo de tipos escriturarios» (Fradejas Rueda, 19/12/2023).

²² En la descripción del modelo aparece la siguiente información: «Coloso Español is a versatile AI model in Transkribus designed to transcribe a wide array of Spanish texts, from medieval manuscripts to 20th-century documents». Además, se sostiene que pronto se incorporarán más detalles y se publicará un estudio dedicado. Los colaboradores expresamente citados que participaron en la constitución del modelo son: Stefano Bazzaco, Alba Comino, Andrés Echavarría Peláez, José Manuel Fradejas Rueda, Francisco Gago Jover, Raquel Liceras-Garrido, Patricia Murrieta-Flores, Humberto Olea Montero, Rocío Ortuño Casanova, Fernando J. Pancorbo, Milena Peralta Friedburg, Eva Sánchez-Salido, Rodrigo Vega Sánchez, Juan Carlos Vallejo Velásquez y Ezequiel Villani (<https://readcoop.eu/model/coloso-espanol/>, cons. 27/12/2023).

²³ Lo que se sostiene está basado en dos observaciones principales. En primer lugar, un examen más detallado de los CER, que son mayores en el set de entrenamiento que en el de validación, lleva a imaginar que existe una variabilidad interna del modelo que la máquina gestiona con dificultades:

búsqueda difusa con herramientas de *Keyword Spotting* (KWS), quizás serían suficientes modelos de este tipo, porque el grado de aproximación del texto transcrito a las realizaciones gráficas en la página parece bastante adecuado para que se puedan sacar algunas informaciones pertinentes, aunque no certeras. Por otro lado, desde el punto de vista de la creación de ediciones que respetan criterios filológicos más estrictos, parecen herramientas a esta altura todavía inservibles. Sobre este asunto, por lo tanto, concordamos con Fradejas Rueda, quien sostiene que «la creación de macromodelos es una idea muy interesante, [...] debemos generarlos, pero no se pueden llevar a cabo por acumulación de *growth truth sets* diseñados para un arco temporal tan amplio [...], ni con criterios de transcripción absolutamente antagónicos» (https://7partidas.hypotheses.org/11531#footnote_6_11531, cons. 27/12/2023). Al contrario, sería preferible que se crearan macromodelos menos extendidos que unificaran modelos parciales basados en tipos escriturarios de un periodo determinado y que siguieran unos mismos criterios, reduciendo al mínimo el ruido generado por la abundante variación interna.

6. CONCLUSIONES

Para concluir, resumimos los puntos principales tratados en el presente trabajo y sugerimos unas posibles pistas futuras.

Al principio, señalamos cómo el reconocimiento de textos impresos y manuscritos es un contexto en creciente evolución que se apoya en las recientes innovaciones tecnológicas. En particular, apuntamos a la introducción de sistemas de HTR en distintos ámbitos de la gestión de la información, subrayando el interés que suscitan herramientas como Transkribus y eScriptorium, ambientes virtuales colaborativos que presentan una interfaz de fácil alcance para todo tipo de usuario y que prometen obtener resultados de transcripción automática hace unas décadas impensables. Al respecto, se valoraron los principales recursos disponibles en el campo de la transcripción automática de impresos antiguos y manuscritos, fijando la atención en los modelos mixtos, es decir, herramientas de HTR de largo alcance basadas en *datasets* heterogéneos que consienten la interpretación de nuevos documentos que no se incluyeron para el aprendizaje de la máquina.

En el caso de los modelos de reconocimiento para impresos antiguos, detallamos la presencia de modelos mixtos de HTR ya disponibles en la plataforma Transkribus, poniendo el acento prioritariamente en los procesos que llevaron a la confección

probablemente al tomar otros materiales del mismo conjunto para la validación se asistiría a una alteración significativa de las tasas de error. En segundo lugar, al poner a prueba de forma somera el modelo con unos textos de nuestras colecciones registramos que los resultados entremezclan criterios de transcripción, a veces proponiendo lecturas muy conservadoras, otras normalizando, sin que se pueda detectar una sistematicidad en las decisiones tomadas. Lo mismo señala Fradejas Rueda a la hora de trabajar con sus textos de interés (https://7partidas.hypotheses.org/11531#footnote_6_11531, cons. 27/12/2023).

de los mismos con la intención de establecer unas buenas prácticas para su futura explotación y difusión. Con respecto a la transcripción automática de documentos manuscritos, por otra parte, señalamos los problemas que todavía afectan a esta área de investigación, que necesariamente requiere perfeccionarse para proporcionar unos resultados asequibles, sobre todo en el caso de los estudios filológicos y la creación de ediciones digitales académicas.

La bibliografía al respecto está experimentando un progresivo aumento y apunta a una mayor interacción entre humanistas digitales, especialistas informáticos e instituciones de la memoria.

Por su parte, los primeros deben adoptar buenas prácticas en la preparación de los *datasets* para el entrenamiento del modelo, favoreciendo transcripciones basadas en criterios comunes y difundiendo en formatos, como XML o ALTO (Analyzed Layout and Text Object), que aseguran la sostenibilidad del trabajo llevado a cabo. Al respecto, la iniciativa HTR United (<https://htr-united.github.io/>, cons. 27/12/2023), proyecto que reúne los *dataset* ya al alcance de la comunidad científica acompañados de metadatos estándar, es de sumo interés porque consiste en entretener y consolidar los esfuerzos de distintos investigadores para alcanzar objetivos colectivos.

Los técnicos informáticos, por otra parte, auxilian la investigación, proveyendo no solamente infraestructuras de colaboración adecuadas e interfaces que no requieren un alto nivel de competencias informáticas, sino también explorando nuevas vías de creación de datos fiables y explotación de los resultados obtenidos. De hecho, existen pocas indicaciones acerca de la cantidad mínima de texto transcrito para crear modelos consistentes, por ejemplo, pasando por la agrupación y jerarquización de los materiales ya al alcance (Perdiki, 2023). Del mismo modo, en lo que atiene a la postcorrección, ha sido observado que para los textos sacados con sistemas de HTR se requieren algoritmos diferentes del OCR, puesto que los errores derivados de la aplicación de ambos son de naturaleza distinta y el riesgo es de generar ulterior ruido en las transcripciones finales (Capurro *et al.*, 2023: 7490); no obstante, la experimentación en este campo sigue siendo escasa²⁴.

Bibliotecarios y archiveros, finalmente, están fuertemente implicados en el proceso, ya que guardan el potencial de extender la variedad de los materiales digitalizados asegurando que la pluralidad de voces que puebla sus colecciones sea preservada. Para alcanzar este objetivo y con ello beneficiar de la tecnología HTR promoviendo la búsqueda eficiente de información, Terras (2022b: 144) sugiere que las instituciones de la memoria tienen que adoptar recursos para el reconocimiento de textos en la misma fase de digitalización de los materiales archivados, buscando vías para garantizar la exportación de los datos de modo sostenible, procesable y reutilizable. En concreto, falta todavía una denominación estándar que permita

²⁴ Las publicaciones de mayor interés al respecto son Neto *et al.* (2020) y Pavlopoulos *et al.* (2022).

clasificar y administrar los contenidos generados con HTR, para los cuales se necesita la producción de identificadores persistentes y la conexión con infraestructuras y *frameworks* de difusión como IIIF que soporten la interoperabilidad entre distintos repositorios (<https://iiif.io/>).

Solo reforzada por estos tipos de sinergias la tecnología HTR podrá ser sistemáticamente empleada para la producción de *source data*, con recaídas en la creación de ediciones digitales académicas, análisis de carácter lingüístico y semántico, o la búsqueda de información, en conjunción con la detección de palabras clave y la inclusión en los Content Management Systems (CMS) de las extensiones digitales de bibliotecas y archivos. De esta manera se podrán garantizar la correcta preservación y circulación de nuestro patrimonio documental, favoreciendo un libre y sostenible proceso de generación de conocimiento.

FINANCIACIÓN

El presente trabajo se ha desarrollado en el marco de las actividades del Progetto Mambrino (Università di Verona), en particular el PRIN «Mapping Chivalry: Spanish Romances of Chivalry from Renaissance to 21st Century. A digital approach» (2018-2023, prot. 2017JA5XAR), del PRIN 2022 PNRR «The digital catalogue of Spanish epic chivalric poems of the 16th and 17th centuries: texts, paratexts and socio-literary networks (an interdisciplinary approach)», del Progetto di Eccellenza «Inclusive Humanities. Prospettive di sviluppo nella ricerca e nella didattica delle lingue e letterature straniere» (2023-2027) y del centro DAIH (Digital Arena for Inclusive Humanities).

REFERENCIAS

- Allés Torrent, S. (2020). Crítica textual y edición digital o ¿dónde está la crítica en las ediciones digitales?. *Studia Aurea: revista de literatura española y teoría literaria del Renacimiento y Siglo de Oro*, 14, 63-98. <https://doi.org/10.5565/rev/studiaaurea.395>
- Alvite-Díez, M. L. y Barrionuevo, L. (2020). Confluence between library and information science and digital humanities in Spain. Methodologies, standards and collections. *The Journal of Documentation*, 77(1), 41-68. <https://doi.org/10.1108/JD-02-2020-0030>
- Alvite-Díez, M. L. y Rojas-Castro, A. (2022). Ediciones digitales académicas: Concepto, estándares de calidad y software de publicación. *El Profesional de la Información*, 31(2), 1-19. <https://doi.org/10.3145/epi.2022.mar.16>
- Ball, R. y Parker, G. (Eds.). (2014). *Cómo ser rey. Instrucciones del emperador Carlos V a su hijo Felipe. Mayo de 1543*. CSA-The Hispanic Society of America.
- Bazzaco, S. (2018). El Progetto Mambrino y las tecnologías OCR: estado de la cuestión. *Historias Fingidas*, (6), 257-272. <https://doi.org/10.13136/2284-2667/89>
- Bazzaco, S. (2020). 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), 534-561. <https://www.janusdigital.es/articulo.htm?id=160>

- Bazzaco, S., Jiménez Ruiz, A. M., Torralba Ruberte, A. y Martín Molares, M. (2022). Sistemas de reconocimiento de textos e impresos hispánicos de la Edad Moderna. La creación de unos modelos de HTR para la transcripción automatizada de documentos en gótica y redonda (s. XV-XVII). *Historias Fingidas*, (Número Especial 1), 67-125. <https://doi.org/10.13136/2284-2667/1190>
- Bermúdez Carreño, J. (2023). Inteligencia artificial para la transcripción de letra itálica española del siglo XVIII: Transkribus como herramienta para las humanidades digitales. *Revista De Humanidades Digitales*, 8, 109-127. <https://doi.org/10.5944/rhd.vol.8.2023.38111>
- Capurro, C., Provatorova, V. y Kanoulas, E. (2023). Experimenting with Training a Neural Network in Transkribus to Recognise Text in a Multilingual and Multi-Author Manuscript Collection. *Heritage*, 6(12), 7482-7494. <https://doi.org/10.3390/heritage6120392>
- Cordell, R. y Smith, D. (2018). *A Research Agenda for Historical and Multilingual Optical Character Recognition*. Northeastern University Library. <http://hdl.handle.net/2047/D20297452>
- Cuéllar, Á. (2023). La Inteligencia Artificial al rescate del Siglo de Oro. Transcripción y modernización automática de mil trescientos impresos y manuscritos teatrales. *Hipogrifo. Revista de literatura y cultura del Siglo de Oro*, n(1), 101-115. <https://doi.org/10.13035/H.2023.11.01.08>
- Firmani, D., Maiorino, M., Merialdo, P. y Nieddu, E. (2018). Towards Knowledge Discovery from the Vatican Secret Archives. In Codice Ratio - Episode 1: Machine Transcription of the Manuscripts. En Association for Computing Machinery (Ed.), *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 263-272). Association for Computing Machinery. <https://doi.org/10.1145/3219819.3219879>
- Fradejas Rueda, J. M. (2022). De editor analógico a editor digital. *Historias Fingidas*, (Número Especial 1), 39-65. <https://doi.org/10.13136/2284-2667/1108>
- García-Reidy, A. (2019). Deconstructing the Authorship of Siempre ayuda la verdad: A Play by Lope de Vega?. *Neophilologus*, 103(4), 493-510. <https://doi.org/10.1007/s11061-019-09607-8>
- Gille Levenson, M. (2023). Towards a general open dataset and models for late medieval Castilian text recognition (HTR/OCR). *Journal of Data Mining and Digital Humanities. Special Issue: Historical documents and automatic text recognition*. <https://doi.org/10.46298/jdmdh.10416>
- Hodel, T., Schoch, D., Schneider, C. y Purcell, J. (2021). General Models for Handwritten Text Recognition: Feasibility and State-of-the Art. German Kurrent as an Example. *Journal of Open Humanities Data*, 7(13), 1-10. <https://doi.org/10.5334/johd.46>
- Kroll, S. y Sanz-Lázaro, F. (2022). Romances teatrales entre Mira de Amescua, Calderón y Lope, ritmo, asonancia y cuestiones de autoría. *Revista de Humanidades Digitales*, 7, 1-18. <https://doi.org/10.5944/rhd.vol.7.2022.31620>
- Liceras Garrido, R., Comino, A. y Murrieta Flores, P. (2022). Mujeres en el Catálogo Monumental de España: Discursos arqueológicos sobre Prehistoria y Edad del Hierro en las provincias de Ávila, Soria y Burgos. *Complutum*, 33(1), 269-288. <https://doi.org/10.5209/cmpl.80895>
- Mancinelli, T. (2016). 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), 255-260. <https://doi.org/10.13136/2284-2667/65>

- Menta, A., Sánchez-Salido, E. y García-Serrano, A. (2022). Transcripción de periódicos históricos: Aproximación CLARA-HD. En M. Á. Alonso, M. Alonso-Ramos, C. Gómez Rodríguez, D. Vilares Calvo y J. Vilares (Eds.), *Proceedings of the Annual Conference of the Spanish Association for Natural Language Processing 2022: Projects and Demonstrations SEPLN-PD 2022*. (pp. 70-74). Universidade da Coruña y CITIC, LYS Research Group.
- Mühlberger, G., Seaward, L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S., Gatos, B., Greinöcker, A., Grüning, T., Hackl, G., Haukkoara, V., Heyer, G., Hirvonen, L., Hodel, T., Jokinen, M., ... Zagoris, K. (2019). Transforming scholarship in the archives through Handwritten Text Recognition. Transkribus as a case study. *Journal of Documentation - Emerald Publishing*, 75(5), 954-976. <https://doi.org/10.1108/JD-07-2018-0114>
- Neto, A. F. de S., Bezerra, B. L. D. y Toselli, A. H. (2020). Towards the natural language processing as spelling correction for offline handwritten text recognition systems. *Applied Sciences*, 10(21), 7711. <https://doi.org/10.3390/app10217711>
- Pavlopoulos, J., Kougia, V., Platanou, P., Shabalin, S., Liagkou, K., Papadatos, E., Essler, H., Camps, J. B. y Fischer, F. (2022). Error Correcting HTRed Byzantine Text. *HTREC*, 1-15. <https://doi.org/10.21203/rs.3.rs-2921088/v1>
- Perdiki, E. (2023). Preparing Big Manuscript Data for Hierarchical Clustering with Minimal HTR Training. *Journal of Data Mining and Digital Humanities. Special Issue: Historical documents and automatic text recognition*. <https://doi.org/10.46298/jdmhdh.10419>
- Pinche, A. (2023). Generic HTR Models for Medieval Manuscripts. The CREMMALab Project. *Journal of Data Mining and Digital Humanities. Special Issue: Historical documents and automatic text recognition*. <https://doi.org/10.46298/jdmhdh.10252>
- Rabus, A. (2019). Recognizing Handwritten Text in Slavic Manuscripts: A Neural-Network Approach Using Transkribus. *Scripta & e-Scripta*, 19, 9-32.
- Schwarz-Ricci, V. I. (2022). Handwritten Text Recognition per registri notarili (secc. XV-XVI): una sperimentazione. *Umanistica Digitale*, (13), 171-181. <https://doi.org/10.6092/issn.2532-8816/14926>
- Souibgui, M. A., Bensalah, A., Chen, J., Fornés, A. y Waldispühl, M. (2022). A User Perspective on HTR Methods for the Automatic Transcription of Rare Scripts: The Case of Codex Runicus. *Journal on Computing and Cultural Heritage*, 15(4), 1-18. <https://doi.org/10.1145/3519306>
- Terras, M. (2010). The Rise of Digitization: An Overview. En R. Rukowski (Ed.), *Digital Libraries* (pp. 3-20). Sense Publishers.
- Terras, M. (2022a). Inviting AI into the Archives: The Reception of Handwritten Recognition Technology into Historical Manuscript Transcription. En S. Jaillant (Ed.), *Archives, Access and Artificial Intelligence. Working with Born-Digital and Digitized Archival Collections* (pp. 179-204). Verlag - Bielefeld University Press. <https://doi.org/10.14361/9783839455845-008>
- Terras, M. (2022b). The Role of the Library When Computers Can Read: Critically Adopting Handwritten Text Recognition (HTR) Technologies to Support Research. En A. Wheatley y S. Hervieux (Eds.), *The Rise of AI: Implications and Applications of Artificial Intelligence in Academic Libraries* (pp. 137-148). ACRL - Association of College & Research Libraries.
- Weber, A., Ameryan, M., Wolstencroft, K., Stork, L., Heerlien, M. y Schomaker, L. (2018). Towards a Digital Infrastructure for Illustrated Handwritten Archives. En M. Ioannides (Ed.), *Digital Cultural Heritage* (pp. 155-166). Springer. https://doi.org/10.1007/978-3-319-75826-8_13



FACULTAD DE FILOLOGÍA
UNIVERSIDAD DE SEVILLA